PERCEIVED CLIQUE DEFINITION IN EGO CENTERED NETWORKS

BY

CHRISTOPHER MCCARTY

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1992

ACKNOWLEDGMENTS

This dissertation is dedicated to H Russell Bernard and Peter D Killworth who were there for me when I needed them most to my son Sean who is my pride and joy, and to Gary Toups (a promise is a promise)

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

PERCEIVED CLIQUE DEFINITION IN EGO CENTERED NETWORKS

By

Christopher McCarty

December 1992

Chairman H Russell Bernard
Major Department Anthropology

Previous research using ego centered networks is based on the assumption that the explanations for strongly tied cliques or subgroups are obvious and can be used as cues to elicit clique membership Such cues do elicit lists of network alters but the claim that these are cliques which the informant naturally perceives has been unsubstantiated This can only be done by allowing informants to freely define their network and measure network ties independently of any specific cues

Forty seven informants were asked to respond to a computerized interviewing task which lasted from two to four hours Informants free listed 60 people they knew where knowing was loosely defined yet constrained Informants ranked on a scale of 1 to 5 how well they knew each alter They then described how they used this 1 to 5 scale Most time consuming was the provision of a structured set of information about each alter including a textual description of how the informant knew the alter Finally informants were asked to rank on a scale of 0 to 5 how well each unique pair of alters knew one another a total of 1 770 pairs

The final module resulted in 47 adjacency matrices which were analyzed with an overlapping clustering analysis to form cliques which were defined by high interknowing. Using the dossiers from each alter these were coded into types, where a type explained the interknowing. Cliques were presented to a subset of 10 informants along with several randomly generated cliques to test the viability of the clustering algorithm and the coding scheme. In virtually all cases, true cliques were distinguished from phony cliques and the coding was found to be consistent with the informant's explanation. Attributes of these perceived types were then analyzed with respect to attributes of the informant, alters, network structure and cooccurrence with other types.

In an attempt to estimate the error associated with my small nonprobability sample of 47 informants, a telephone survey was implemented to find the distribution of the types of social relations among Floridians. Some specific differences are noted.

TABLE OF CONTENTS

ACKNOWLEDGMENTS iii

ABSTRACT vii

CHAPTERS

1 INTRODUCTION 1

2 A THEORY OF SOCIAL RELATIONS [illegible]

Knowing Versus Social Relations [illegible]
Maintenance of Relations [illegible]

3 SOCIAL NETWORKS [illegible]

Introduction [illegible]
Development of Network Theory 11
Clique Detection 15
Ego-centered Versus Bounded Networks 20
The Status of Social Networks as a Science 24

4 METHODS 26

Computer Versus Paper Formats 26
The Sample 26
Data Gathering on Informants and Choosing Alters 28
The Autocorrelation Problem 30
Ranking Alters on Levels of Knowing 31
Alter Detail 37
Pair Rankings 46
Network Dynamics 53
Summary of Alters 54

5 GROUPING ANALYSIS 56

Clustering Procedure 56
Cluster Types 58
Ego/row Versus Non ego/binary 62
Verification of Clusters 65
Variance Explained 66
Informant Characteristics and Clusters 67
Within Cluster Similarity 69
Correlations 71
Factor Analysis on Cluster Type 73
Overlap 75
Free Listing Order 79
Recalculation of Clusters 82
The Meaning of Clusters 84
Telephone Survey 85
Network Density 91
Network Degree 94

6 CASE STUDIES 97

Introduction [illegible]
Steve [illegible]
Betty [illegible]
Tony and Mary [illegible]
Jennifer [illegible]
Selected MDS Plots [illegible]

7 DISCUSSION AND CONCLUSIONS [illegible]

Experimental Method [illegible]
Knowing [illegible]
The Social Relation [illegible]
Clique Definition [illegible]
Conclusion [illegible]

APPENDIX 138

REFERENCES 139

BIOGRAPHICAL SKETCH 144

CHAPTER 1
INTRODUCTION

Imagine a list that contains all of the people you have known in your life. If it were possible to dredge up such a list, some people on the list would not know others on the list, while some people would know others. Is it likely that these social relations occur randomly, or are there clear explanations which make it possible to predict them?

Who you know, and how you know them, is a function of where you are, what you do, your likes and dislikes, and almost certainly characteristics such as your gender, race and marital status. Therefore, the groupings, or subnetworks, of your global network are probably subject to the same phenomena. Who your relations know among your relations, and how they know them, are a function of these same factors.

Furthermore, your position and relation to these groupings is not static. Time and life changes affect social relations. It is likely that each person has a global network of active relations, or relations which comprise the majority of their interaction and behavioral prediction, as well as a cumulative global network of all relations ever held. For an adult villager in the mountains of China, the list of current network actors may be very similar to his or her list of network actors ever known. People in such villages may grow old maintaining mostly the same set of relations they accumulated by age four. Additions to their network are most likely due to the births of potential members in the village. For mid level executives in national American companies, where transfers and electronic communication are an ordinary part of life, the list of actors ever known is almost certainly much larger than their currently maintained list of actors.

In this thesis I will present the results of an experiment into knowing, and the definition of perceived subnetworks based on a given definition. The approach is presented as a return to the substance that early anthropologists studying network phenomena chose to examine utilizing some of the methodological advancements that are the hallmark of contemporary network research. This work is a contribution to the field of social networks in identifying categories of subnetworks that are meaningful to informants. On a more general level it is suggestive of the dynamics of personal networks and the critical roles played by location and other variables in their formation.

In the next chapter a theory about social relations serves to outline my views as to how social relations are formed and maintained. Chapter 3 is a review of the history of social networks, with a particular focus on clique finding algorithms. Chapter 4 describes my methods, sample and some results relating to the ways in which informants used the concept of 'knowing'. Chapter 5 is the heart of the thesis, presenting all of the analysis based upon the results of the overlapping clustering analysis using the ADCLUS model, as well as the results of a telephone survey method used to verify the results from extended interviews with 47 informants. The sixth chapter consists of five case studies which help illustrate many of the concepts presented in Chapter 5. In the final chapter are discussions of the experimental procedure and conclusions.

CHAPTER 2
A THEORY OF SOCIAL RELATIONS

Knowing Versus Social Relations

Babbie (1989) points out that the ambiguity of concepts in social science lead to an ironical cycle. For example, we observe behaviors such as people visiting one another or sharing life experiences, and we perceive that these are related behaviors which we describe under the umbrella term 'knowing'. As we talk about these behaviors using the abbreviation 'knowing', we come to think it truly exists, rather than being a summary of many behaviors. Finally, we take to task the study of 'knowing' as a thing that exists or, even worse, use the word as a stimulus to evoke a definitive behavior across informants. Is it any wonder that estimators of network size which use 'knowing' as their stimulus for network tie verification vary wildly (Pool and Kochen 1978, Burt, 1982, Bernard et al., 1987, Bernard et al., 1989, Freeman and Thompson, 1989)?

When people talk about knowing someone, they are actually using an abstraction (or variant) of the concept of social relations. Unfortunately, knowing is a subjective term that is far from a standardized concept in social science. The limits of the concept of knowing are made abundantly clear when other languages and cultures are examined. Excluding the ambiguity of the term in its use among speakers of English, there is not a clear equivalent for 'knowing' in many languages. It is a tenuous term of questionable value as a central concept to network theory. We use the term because our informants have an understanding for it which we assume approximates the relation which interests us. That this definition varies across informants one rationale behind this thesis.

A reasonable alternative to the concept of knowing is the social relation a term which has been used time and again by social scientists since Emile Durkheim This concept is descriptive and general enough to include all specific relational concepts such as knowing, closeness or even love As social network analysts we are interested in the structure of social relations their content, and how these phenomena interact in a lawlike manner

What, then, is a social relation? How do we decide a relation exists and how is it best described? It does not take long to realize that this is critical to the size and content of both the global network and any subgroupings Defining a social relation as anyone a person has ever talked to yields a much larger global network than defining it as those who a person talks to about important matters A general construct for social relations is much preferable to the infinite operational definitions which breed disagreement and doubt among researchers

Understanding social relations is more sensible if thought of in terms of behavior prediction or expectation This is a set of concrete and potentially measurable phenomena which can eventually be standardized

Imagine two people, a boss and her employee who meet for the first time To have assumed these roles these two actors must have accumulated years of life experiences which have conditioned expectations as to how certain people act both personally and in institutionally prescribed ways Thus the employee will predict a boss to act in a particular way to certain behaviors Showing up at noon drunk will most likely draw an unfavorable response from the boss unless personal experience dictates otherwise There are certain behaviors exhibited by the boss that the employee's prior experience will enable him to predict Others such as how the boss will react to particular kinds of clothes, may require at least one personal experience to form a basis for prediction.

One aspect of the social relation is a set of predictions that one person has of another person's behavior. Behavior can be the passage of information, provision of a resource, or causative action, all real or potential. The descriptions of knowing explored in this thesis are all derived from this principle. To say 'I know him, he is my brother' is a relation one which describes predictions. To say 'I know her, I see her a couple of times a week' suggests more accurate or more specific predictions about behavior than the statement 'I know her, I see her once a year.'

Social relations are often asymmetric. At the extreme there is a social relation between myself and the Queen of England[1]. I can make some crude predictions about her behavior from information contained in the British tabloids and from my limited knowledge of royal protocol. But the Queen can make no predictions about my behavior. It is more accurate to say I 'know of' the Queen, specifically because she does not know of my existence and thus cannot and will not make predictions about my behavior. *Knowing of* someone implies a completely asymmetric or one sided social relation, where *knowing* someone implies a partially asymmetric or symmetric social relation.

Consider next my relation to the Dean of my College. There is much more potential for the Dean's behavior to affect me than mine his. In fact, on many occasions I have predicted his behavior, but he rarely makes predictions of mine.

My department head and I are very much affected by each others' behavior. I often make predictions about her behavior, as she does about mine. Although these predictions may be partly based on situational rules there is a lot of flexibility in these rules and therefore variability in the set of possible behaviors we each can emit

---

[1] Mythical figures, such as Santa Claus, are clearly unfounded. And social relations with religious figures, such as Jesus, are, like Socrates, excluded. In past experience, many religious informants have insisted that they have a relationship with Jesus. I argue that it is not a social relation.

So social relations can be viewed as the prediction of behavior accurate or not and its intensity measured by the amount of behavioral consequence the actors have on one another and their accuracy in prediction This view provides a consistent framework to anchor operational constructs It excludes people like Michael Jackson about whom I know something but who knows nothing of me and does not affect me[2] It includes store clerks with whom I have ephemeral contact because we must make some predictions about one another's behavior It also includes people I have not met personally if we know of one another through a mutual third party These indirect relations can become direct relations and are the source of much dynamism in social networks A sociometric graph would show these people on the fringes of each others ego centered networks

A social relation exists between two people if at least one and frequently both actor(s) can make specific behavioral predictions about the other and that there is some personal consequence of the actor's behavior no matter how minute these consequences for the other

Maintenance of Relations

So far social relations have been defined in a static sense But social relations are affected by the life changes that go with time Thus I have known people in the past whom I do not know now And I know people now whom I will not know in the future

Maintenance of a social relation depends on 1) my ability to make accurate predictions about a person's behavior and 2) the degree of consequence that person's behavior has on me

---

[2] This definition of a social relation ultimately rests on what are and are not considered behaviors For instance are behaviors necessarily exactly observable phenomena or is an attitude a behavior?

Physical proximity is the primary foundation for behavioral prediction[3] Evidence for this comes from a series of experiments conducted by Bernard, Killworth and McCarty (1982), Killworth, Bernard and McCarty (1984) and Bernard et al (1988) where location was overwhelmingly cited as the primary reason for making a choice in the first link of a social chain from the informant to some random target In fact, most serially proximate relations are based on physical proximity either now or in the past, of the links (dyads)

Serially proximate relations are somehow dependent on the relations of other persons to be maintained Some of the best examples are work relations, where the foundation for the relation is the collective goal of other employees to perform their work or the "via" where the continued relation of a third party is necessary to maintain the relation

Thus the primary factor affecting the maintenance of a serial relation assuming that we want to maintain it is change in physical proximity or secondarily a change in serial proximity In my case most of my serial relations will disappear given a move to Alaska[4] Changing jobs even moving to another department will have a dramatic effect on my current global network And too many people have experienced the effects of divorce on their relations with relatives and friends

In general predicting a person's behavior depends on how much interaction we have or have had personally or through others When a life change like relocation or divorce occurs one must reassess how much time or other resources must be expended in order to maintain each social relation and whether the expenditure is worth the rewards

---

[3] The argument could be made that media, such as telephones and fax provide a type of physical proximity All media enhanced social relations originally arise from physical or serial proximity

[4] Recently, a friend and colleague of mine who is about to receive his doctorate took a position in France and was due to leave soon He commented that people he has known for the past four years reacted to him in strange ways His description of their behavior was something akin to *grieving* They are not grieving for him he is obviously moving to better circumstances They are grieving the loss of the relation which in most cases will disappear

Some relations of course are so conditioned that there is little choice of whether to maintain them Most parent-child relations fall into this category (long distance telephone companies depend on it) As will be seen from this experiment this still holds in today's mobile and technology based society

CHAPTER 3
SOCIAL NETWORKS

Introduction

The Holy Grail of behavioral science is the development of a body of theory and, in turn, research strategies, which share the explanatory power, validity, replicability and application which is now accepted as common in the natural sciences. Our collective progress in achieving this is debatable. If volume were any indication of the development of a science, the study of human behavior would have overtaken natural science decades ago. Many social scientists are embarrassed by comparisons of their work to physics and chemistry. Others revel in the perceived imbalance in this comparison, striving to treat humans as something different than the stuff of science.

Perhaps the biggest problem in behavioral science has been our preoccupation with drawing territorial boundaries. This has caused infighting within disciplines, leading to the frustration of many academics who see value in multi disciplinary frameworks. Oddly, this turf war seems unique to social science. Imagine a chemist and a physicist arguing whether the study of the behavior of a particle were a legitimate pursuit for one or the other, rather than focusing on the methods and results of their studies.

Despite our feelings of inferiority, calls for preserving the humanity in the study of humans, and fruitless bickering over who may study which phenomena, there have been enormous advances in the study of human behavior over the past decade. Several disciplines have developed tried and true applications of theory and experimentation. They have contributed to treatment of crippling anxiety, to nurturing tolerance of others and to the conduct of economic exchange.

But like any science certain assumptions must be met to bring about advancements which all agree to be progressive Popper (1953) refers to the falsifiability of scientific propositions inductive propositions should be considered suspect and data collection should be done in order to test theories which put forth risky propositions These are propositions which upon being tested may be negated as easily as they are substantiated. Large scale theory in behavioral science (such as it is) is almost exclusively inductive and non refutable Smaller scale theories do better but are often the result of induction

Kuhn (1962) describes science in terms of paradigms A paradigm is a set of achievements which has the power to explain previously noted phenomena A paradigm is also open ended in that it suggests experiments for the testing of new phenomena A shift occurs when phenomena that cannot be explained by the prevailing paradigm are explained by a new paradigm Paradigmatic propositions must be verifiable and testable If they are not then conflict will not result in newer and better paradigms Scientists who disagree on a conclusion must have some mechanism to trace their reasoning back to a point where their theories clash so that one view can be shown in error Most behavioral science lacks the mechanism for two scientists to debate and for one to come away admitting that he or she was wrong One can probably count on one hand the number of times the author of a behavioral theory admitted that the theory was in error and was supplanted by another

Bernard and Killworth (1979) present reasons why we have yet to develop a "social physics " as suggested by August Comte (The Positive Philosophy 1896) One of the core reasons is our propensity to analyze only the detail of human behavior rather than the larger mean behaviors from which lawlike relationships arise This is not to say that the recording of detail is unproductive It is the lack of testable hypotheses which is the base of behavioral science and testable hypotheses depend on succinct and operationally sound description. Social network analysis is a promising approach for presenting hypotheses in a testable and replicable manner.

Development of Social Network Theory

Specific to social science, in contrast to other areas of behavioral science, is the focus on social relations. Emile Durkheim was an important figure in establishing this agenda by emphasizing the relation of individuals to the social group (1933, 1938). Drawing heavily from Comte and Saint Simon, Durkheim turned from primarily biological reductionist and evolutionary theories of society to an explanation that attributed more to the effect of the group on the individual. His views, that 'social facts' were part of an independent reality which constituted a group mind, ultimately led to the birth of the French structural school, identified today with Levi-Strauss (Harris, 1968, pg. 472). Despite these views, which are anathema to many contemporary positivistic social scientists, Durkheim pointed to the relation as the substance of social science.

Radcliffe-Brown was one of the first to allude to social structure as a network of social relations. Drawing heavily from Durkheim, Radcliffe-Brown strove to develop laws of social relationships. He and a generation of students of both anthropology and sociology applied the structural functional approach in their comparative study of social phenomena world wide. In his presidential address to the Royal Anthropological Institute[5] Radcliffe-Brown laid out the fundamentals of his views which were applied in earlier writings (1931, 1933).

In this address, Radcliffe-Brown set the agenda of social anthropology, the study of the network of social relations. He proclaimed all social relations to be the product of this network which is the building block of "social structure." Social structure was a relatively steady set of relations whose component parts social relations, were dynamic exhibiting renewal and attrition To understand the components comparisons between societies were necessary to recognize varying structural patterns[8]

---

[5] Later published as On Social Structure (1940).

Although Radcliffe Brown's use of the concept of the social network remained a metaphor he set the stage for future social scientists to develop it further The tradition of British structuralism gave rise to the description of anthropological phenomena in a common conceptual framework that of individuals connected by relations which provided a channel for social effect (Nadel 1957) Early network studies had as their focus typical anthropological and sociological topics

Barnes (1954) was among the first anthropologists to apply the network concept to field data His description of the social structure of a Norwegian fishing village was grounded in the thesis that three 'fields' of social relations existed the locally defined the industrial and a pervasive system of friends and kin which fused them The first two were seemingly distinct fields One was fixed by locality and traditions of land tenure that encouraged extended families to reside in the same place over decades even centuries The industrial particularly the fishing industry depended on the dynamic interchange of members from one boat to another These were bridged by the informal network of friends and kin demonstrating an exclusively network function

Elizabeth Bott (1955 1957) is recognized as one of the first to apply the social network framework to a research problem in specific terms Bott chose to explain the variation in conjugal roles in selected English families by their social milieu A particular family's social network was found to be highly influential on the division of labor and the patterns of leisure activity they exhibited in their own household In demonstrating this Bott employed, in a mostly qualitative sense concepts such as network density and connectedness which are now operationally defined using statistics and graph theory

---

[8] This view of social structure being largely a steady state which affects and explains human behavior predates the position of contemporary structuralists (see Wellman and Berkowitz 1988) Radical structuralists attribute a large portion of human social behavior to the constraints and stimuli of social structure Although they do not say that social structure exists as a phenomenon in and of itself (see Kroeber 1917) they come very close

This shift of some anthropologists towards structural explanations accelerated following World War II The increasing migration of villagers to urban centers in Africa and their switch from agrarian to an industrial base called for adaptation strategies that were inherently network oriented These strategies were noticed by British structural anthropologists in the 1960s (Mayer 1961 Mitchell 1969 Kapferer 1969 Epstein 1969)

Wolfe (1978) suggests that until anthropologists studied urbanization in Africa they were not faced with network phenomena in need of explanation and thus the study of networks as a discipline did not progress For Wolfe anthropologists were predisposed to recognize network phenomena operating in urban situations given their prior experience with kin systems in rural areas Still with little mathematical expertise networks remained a conceptual framework for explaining social phenomena ethnographically

A lesser known volume of research which applied the network model comes from rural sociology Although this work is rarely cited in the typical treatise on the development of social networks certain works stand out as clear examples of the way in which the model is applied today (see Loomis 1944 1950) Although the analytical methods have advanced many of the questions which we focus on today were addressed then

Many social scientists who do not employ the network approach characterize it as a mystifying bag of methodological tricks which one normally expects from a statistician or mathematician Without question this is the trademark of network research The domination of networks by quantitative methods stems from the influence of Simmel (1955) and an already thriving statistical tradition in American sociology which by 1965 had picked up the network metaphor (Wellman and Berkowitz 1988) Simmel again drawing from Durkheim suggested that social structure provides the constraints on individual behavior and relations between individuals This led to a concerted effort to understand and describe social structure so that it might later explain micro level behavior These hopes were bolstered by the experience of the British structuralists who had already demonstrated quite convincingly that the network metaphor was a helpful framework for explaining social phenomena

Early efforts in using diagrams to illustrate social relations (Moreno 1934 Coleman 1961) and the diffusion of information (Rapoport 1979 Rogers and Kincaid 1981) led to the application of graph theory to network structures (Harary et al. 1965 Holland and Leinhardt 1970) The graph theoretic approach was truly the dawn of the methodological boom in network analysis Finally networks could be represented in such a way that they could be reduced to algebraic expressions akin to the representations of the stars by astrophysicists This meant that given certain assumptions these expressions could be manipulated to discover relations which were perhaps not obvious Although this elevated the study of social relations to the methodological rigor of physics many of the initial assumptions and conceptualizations were so far removed from reality that they were largely irrelevant[7]

---

[7] Strauss and Freeman (1989) quotes Massarik and Ratoosh (1965 pg 17) in describing the "seductiveness of mathematics" in behavioral science

> There are some reasonable and appropriate assumptions about human behavior that at the outset guide the development of a given mathematical model The basic propositions are stated abstractly and from them a number of consequences quickly follow These too are amenable to succinct and precise

Later developments led to the use of statistical routines for the analysis of matrices which were flexible in representing a variety of network concepts (Forsyth and Katz 1946, Festinger 1949, Luce and Perry 1949, Alba 1973). Where graph theory is ideal for the representation of small groups and for mathematical modelling, matrix oriented methods are better suited for both etic and emic data collected from larger groups. In particular, the detection of cliques within large groups of network actors became the forte of those using the matrix approach. Clique detection became an important part of this research.

Let me make it clear that social network analysis is not a theory. Rather it is a research strategy or a substantive focus where the propositions tend to be testable. Although some low level theories arose from this strategy/focus (see Granovetter's theory of "the strength of weak ties" 1973), none offers an explanation of large chunks of social phenomena. Social networks is so far a research tradition with the potential to offer a paradigm and propositions which are falsifiable and refutable. Its greatest contribution to date is in developing standard measures of social phenomena upon which all network analysts can and do agree (e.g. density, connectedness, betweenness).

Clique Detection

Freeman (1991) points to the preoccupation sociologists have had with the concept of "group." Groups definition is central to the fundamental theories of much of traditional sociology. Although sociologists and anthropologists have expressed what Freeman calls a

---

notation. Soon a network of rigorous propositions is derived and mathematically manipulated. However, as the tower of derived propositions is built higher and higher, the danger increases that knowledge about human behavior is left further and further behind. The model itself becomes a soul satisfying, plausible, and conceptually elegant end product "sensitizing" view of groups (that is an intuitive feel for what groups are) operationally defining them has proved a difficult task

Groups like cliques are collections of individuals who are connected by some form of affiliation Unlike cliques groups are by definition non overlapping that is members of a set can be partitioned into two or more mutually exhaustive clusters[5] Therefore cliques represent clusters of affiliations within a set without the necessary constraint of non overlapping members

The historical development of clique-finding algorithms has two distinct sides One side the more debated is the development of the computational algorithm which takes data in binary or valued directed or undirected form and constructs subgroupings or cliques Less discussed but in my opinion just as important is the method for generating the data The stimulus presented to the informant or the rule which determines the notation of an etically defined tie (dyad) are the input for the formal analysis Different stimuli or rules can generate violently different results In fact this issue is part of the justification for my research

Until the middle 1970s the development of clique finding models and algorithms was sporadic With the advent of larger and faster computers this research blossomed Now a variety of clique finding techniques exist focusing on different aspects of the clique structure These are not competing algorithms in the search for the "true clique " It is entirely up to the researcher to decide which characteristics should be highlighted And it is up to the reader to decide if these are germane to the explanation of social phenomena

---

[5] The concept of group seems to be honest completely divorced from reality Most efforts to define it such that data fit the model seem forced often requiring heroic assumptions Furthermore it seems that any channels for network effects are strengthened by the existence of overlap and certainly are not necessarily hindered by it Thus my appreciation for the struggle to define groups as nonoverlapping simply because it is intuitive from the traditional sociological view is minimal

Luce and Perry (1949) developed the first operational clique detection procedure.[9] For this routine a clique was defined as a subset of individuals all whom know one another. Being a matrix procedure, the input was a standard adjacency matrix where one cell represented the knowing between two individuals. Luce and Perry limited their analysis to binary data where cells contained a 1 for a tie and a 0 for a nontie. Also, ties were assumed to be symmetric, that is, the upper and lower halves of the matrix were mirror images of one another with zeros down the diagonal. Symmetry implies that person A and person B know each other with equal strength. A tie could be based on a variety of definitions, from strict observation of physical contact to self reports of whether actor A liked actor B.

This definition of cliques was a breakthrough methodologically; still, it was considered problematic by Luce and Perry as well as by those whom they influenced. First, the assumption of a maximal complete subgraph, i.e., all members know one another, is too stringent. A subset of highly affiliated actors where some ties do not exist is easily imagined. Second, the restriction to binary data ignored the variability in tie strength. For instance, defining a tie using length of time actors have known one another, it is obvious that there would be high variability in some cases in the strength of the tie. A third criticism is the assumption of symmetry. Not only might the ties between pairs of actors vary, but individual actors within a pair might give discordant assessments of their tie strength, assuming the definition of the tie allows for this.

In 1950 Luce presented a refinement called the n-clique. Unlike the previous algorithms which detected maximally complete subgraphs, in this implementation he relaxed the assumption that all actors in the clique know one another directly. Instead, a level of indirect ties, n chains, could be established which provided for the existence of some intransitive ties of a rank less than n.

---

[9] Festinger (1949) published a more abstract and descriptive account of this matrix algebra application almost simultaneously with Luce and Perry. In fact, all three authors were involved in the research together at MIT. The idea of representing ties between an established set of actors in matrix form is attributed to Forsyth and Katz, although their analysis was simple and tedious.

Alba (1973) modified the n-clique concept further by requiring the diameter of the resulting clique to be equal to n. By limiting clique membership to those actors who are tied to all others by paths of n points or less, a resulting n-clique can actually contain paths greater than the social distance allowed. Restricting the diameter of the n-clique to n solves this problem. It makes clique computation easy and allowed Alba to write one of the first packaged clique finding programs (COMPLT).

Seidman and Foster (1978) suggested still another variation of the graph theoretic approach to clique detection. The k-plex model is designed to control the reachability within the subgraph by imposing the restriction of minimum star coverage. Thus any member of the subgraph must be connected to n-k other members by length 1, where n is total clique membership and k is a tolerance point selected by the researcher. The resulting subgraph exhibits certain desirable properties which Seidman and Foster see as complementary to other existing clique-finders. Alba's concerns about diameter are implicitly addressed by the model.

Killworth and Bernard (1974) developed CATIJ as an alternative to the graph theoretic based algorithms which had proliferated since Luce-Perry. Despite the use of matrix input, these algorithms depended on definitions which were based on tolerance points of present or absent ties, or transitive and intransitive relations. CATIJ is unique for two reasons, the most widely cited being the use of rank order data as input.[10] Secondly, it is possible to end up with cliques which are not necessarily highly interactive, but rather generally perceived as interactive groups. This is due primarily to the type of prompt used for collecting data and the method of searching alternative paths between alters.

Another path taken towards clique definition was the application of widely used multivariate techniques such as factor and discriminant analysis or multidimensional scaling (MDS) (see Bock and Husain, 1950; Riffenburgh, 1966). Factor analysis and MDS are based upon the notion of underlying dimensions in the data. Factor analysis in particular attempts to explain as much variance as possible with the initial factor, and as much of the remaining variance with subsequent factors. This is not intuitively appealing when viewing cliques as interacting subgroups. MDS is primarily used for graphic representation of groupings, providing no guidelines for creating distinct boundaries around cliques. Lankford (1974) achieved good results from factor analysis compared to MDS, linkage analysis (Mcquitty, 1957) and Hubbell's method of raising a square matrix of weighted links to the nth power (1965).

Peay (1974, 1976) characterizes a set of clique-finders as hierarchical cliquing routines. These methods capitalize on the strength of ties. Early examples of hierarchical algorithms come from Harary et al. (1965) and Doreian (1969). Peay points out that these approaches invoke clique-finders for each level of relationship. Thus, all ties of level 3 are examined for cliques, then level 2, etc. Peay introduces an algorithm that works on minimum tie criteria. This has the advantage of considering the values of more than one level at a time. Peay's method constitutes a type of hierarchical clustering analysis. Many network researchers use various clustering routines as clique finders.

---

[10] Informants in a closed social system are led through a series of sorting routines which result in a matrix where a column represents the ordering for a given informant, all members of the system relative to a question such as "Who do you talk to the most?"

Jain presents a distinction between inclusive and exclusive clustering techniques (1988). Exclusive clustering algorithms, by far the most commonly used, allow objects to belong to only one cluster. Although exclusive techniques are appropriate for many situations it is arguable whether they are the best choice for social network data. Goal orientation has a lot to do with this. Some social scientists suggest that social groups should be defined as mutually exclusive and non-overlapping (Freeman, 1991). From a methodological perspective this makes sense since exclusive groups are more amenable to mathematical modelling. Intuitively though, it is difficult to accept the notion of sub-networks where members can belong to only one group.

Inclusive techniques allow for a different approach, providing for the creation of non-exclusive groups. Arabie, Carroll, DeSarbo and Wind (1981) proposed an overlapping cluster analysis (ADCLUS) that allowed objects to appear in more than one cluster, but eliminated those clusters which did not contribute much to explaining the underlying structure. Ultimately, this method was chosen for its ability to find overlapping clusters (or cliques in this study). The method is available in the Statistical Analysis System (SAS).

Ego-centered Versus Bounded Networks

There are two ways to view the study of networks, or more specifically, what defines a network. These are the etic approach where networks are defined objectively, in this case by society (consensus), and the emic approach where networks are defined subjectively, by the informant.

The etic approach has as its focus the study of bounded, socially defined networks. These are groups which are somehow socially labelled and are expected, by some consensus of members of that society, to exhibit interactions specific to that group. Bounded networks can be clubs churches prison cells or students in a classroom Geographic boundaries can also define bounded networks such as villages or even towns We can talk about a network existing in the United States as a whole although most people would have reservations about the low density and connectedness of this network At some point bounded networks become trivial as a concept

The beauty of bounded networks is that they are perfect for matrix oriented statistical methods which are the foundation of most network methods thus far Assuming that they are of reasonable size bounded networks provide a list of members who can be measured by the researcher and used as a cue for potential informants They are frequently viewed as microcosms of some greater social structure, or at least the elements which are its substance Furthermore bounded networks can be observed and measured over time since the elements are identified and usually expected to be in some predictable location at some point in time

Although they are methodologically elegant the focus on bounded networks suffers from a serious problem How relevant are bounded networks to a person's behavior? Since they are being defined on the basis of methodological convenience they do not necessarily have any great effect on any one member For example the study of church members may reveal high interactions among some members but little among others Since the study is limited to the church membership we are always limited to the effects of that membership on any one member's behavior

It is easy to see that such a view of the social environment is extremely constrained In complex societies such as this country people have social relations with members of several bounded networks What is more some very influential networks might not be easily defined socially, they are non-intuitive to all but a small segment of society such as a group of high school friends

Therefore, limiting our study to bounded networks, though methodologically convenient, may not be relevant. The power of other unmeasured network effects in a person's life could negate the effects of any one measured network.

The emic approach is to define networks from the perspective of each individual. We can think of each person in this world to be the center of a network of social relations. At the extreme, we might even suggest that we are directly or indirectly connected to every other person on this planet, although this has yet to be demonstrated on any functional level.[11]

From a practical standpoint, an emic, or ego-centered network probably has functional boundaries. That is, we should be able to decide, as scientists, upon some measurable criteria for defining the edge of any person's personal network. As was mentioned earlier, how we define the edge will directly affect the size of this network. Using the typical edge-identifiers such as "Who do you know?" or "Who do you talk to?" yield results, but it is unclear what these results are, even when the cues are delivered in a consistent manner. So, there is room for improvement in the way the edge, or boundary, of the personal network is defined.

Assuming we can define ego centered networks empirically, what advantages do they have over bounded networks? Not surprisingly, they are more likely to demonstrate effects on individual behavior than bounded networks. Where bounded networks do not capture the cross sectional nature of an individual by ignoring those networks which are not the one being studied, ego-centered networks by definition include all of these. If we are going to find network effects on individuals in any general sense, it is going to be in the study of ego-centered networks. They are always relevant.

---

[11] Although I may be connected to the villager in the mountains of China by some circuitous route of ten links, it is unlikely that I could have any effect upon his behavior through that route unless I command a certain level of resources. Assuming I *could* have some effect, I am not aware of his particular existence in the first place and would never think to attempt an effect on his behavior.

However, the study of ego centered networks is not problem-free. First, they are not easily measured or observed. Since they are, at least in complex societies, usually composed of memberships to many subnetworks, members may be lured to several physically different locations at any one time. This is usually unpredictable, and therefore makes measurement difficult. So far, we have relied upon the informant to provide what we hope is accurate information about his or her personal network members (alters). Obviously this leaves significant room for error which remains largely unestimated.

Secondly, since each ego centered network is in some way different in both membership and structure, we are faced with the difficulty of comparing physically different networks. Unless we are somehow able to command, in terms of measurement, all of the possible choices for an ego centered network, we are reduced to describing the presence and relations of characteristics to effects in ego centered networks. Those serving life sentences, isolated hunter-gathering bands or perhaps some isolated peasant communities provide some of the few examples of situations where we can be assured of such measurement power.

A third problem concerns the multi network characteristic of any given personal network. Most network researchers talk about personal networks as one large bounded network. That is, they ignore the fact that there are many potential clique effects in any given personal network. When subnetworks are recognized, it is somehow assumed that we already know what they are and how to cue the informant to tell us about them. This is probably erroneous.

My interest in ego-centered networks stems from my firm belief that they will be the venue for the discovery of network effects on individual behavior. In this thesis I am addressing two of the problems listed above. Specifically, I want to contribute to the definition of the edge or boundary of the ego-centered network by exploring the meaning of knowing. More importantly, I will determine the types of subnetworks which informants perceive, independent of the researcher own perceptions.

The Status of Social Networks as a Science

The evolution of any scientific discipline follows the same path from a period of description to one of theory, and ultimately, prediction. In this sense, social networks as a discipline is two-thirds of a science (personal communication, H. Russell Bernard).

Social network research is heavy on description, demonstrating the most elegant and methodologically imaginative attempts to quantify the study of social relations of any of the social sciences. Theoretically the discipline is weak. There are, really, no high level theories comparable to relativity in physics, or structuralism in anthropology. Several interesting and well documented mid-level theories exist, but as yet these have had limited general appeal among other social scientists.

The real problem is that of prediction. What exactly do social networks cause? After all, if they cannot be shown to cause anything then do they really exist? Perhaps they are just interesting abstractions, optical illusions playing on the aspirations of conscientious researchers. Unfortunately this possibility is only too real.

Of all the studies involving social networks, the early studies by anthropologists represent the best efforts to show what networks do. Barnes showed them to mediate limited resources in a small village. Mitchell and others showed them to have made available resources to African villagers migrating to the cities. In most cases, social networks are seen to cause some form of change in the resource structure available to its members compared to the case where members are immersed in an anonymous society. The most fruitful search for network effects will probably derive from this view of the social environment as a resource base.

Not only is it important to show that social networks do indeed cause something, but it has become increasingly important for science to justify itself by addressing practical issues in society as a whole. Although description is important, network analysts must be sure to apply the methodological breakthroughs to the explanation of phenomena of interest to society. Ignoring the demand for application is a sure prescription for the extinction of social networks as a science.

CHAPTER 4
METHODS

Computer Versus Paper Formats.

In this study informants were interviewed using a computerized instrument. Computerized interviewing is becoming increasingly popular over paper and pencil formats in the social sciences (Saris, 1991). Computerization offers immediate coding and data entry, the ability to edit informant responses to a previously established range of answers on closed-ended questions, and as will be seen in this experiment, the ability to preassign questions based upon prior responses. All this is prohibitively expensive when done manually.

On the other hand, computer programs are more complex than paper instruments. This makes them subject to "bugs" in programming and computer hardware failure. Hardware failure is particularly troublesome, not from the standpoint of losing data, which is relatively easy to prevent, but because of the difficulty in reentering a program at the point where the informant was cut off. Still, on the whole, the computerized format was much better for this experiment than paper and pencil would have been.

The Sample

Let me stress up front that this sample was not randomly selected. Given the length of the interview and tedium involved, I decided to use personal acquaintances as some of my informants. Several couples were chosen in the hope of testing for overlap between their responses. Consequently, the results cannot be generalized to any population; they are purely descriptive of the sample.

In all, 47 informants completed the instrument. Table 1 summarizes a variety of their characteristics. Approximately 20 informants were selected among my circle of friends and co workers or their friends The remainder of the sample responded to a newspaper advertisement and were selected in an attempt to maximize a balance based on sex age and race This was clearly not achieved

The sample is heavily biased towards whites This is unfortunate given the interesting results which will be seen from the four black informants A replication of this study using equal numbers of whites and blacks would be useful

Another bias in this sample is the low average age Several informants were students at the university or at a local community college Thus many of them were in their middle to late twenties The fact that several informants were students is shown to have had an effect on the results reported in the following chapter

As an incentive most informants were paid for their participation Informants recruited through friends were paid $25 while the last set of informants recruited from the newspaper, were paid $35 This was thought to have had an effect on the informants obtained through my "snowball sample." However recruitment through the newspaper drew a few informants who were only motivated by the monetary reward Although it was thought that the design of the instrument would make cheating difficult the motivation of some informants to provide useable data was obviously low Again this had noticeable effects on the quality of the data and on my ability to interpret the results It is simply a difficult task to recruit a sample for an instrument which is potentially four hours long

Although there was some typing involved in the response to this instrument informants were not excluded if they had no typing skills In cases where informants could not type I typed for them until they reached a section of the experiment which called for precoded responses The majority of the experiment was precoded and presented no problem to those who could not type

Table 1 Characteristics of Informants

Variable	Category	Frequency	Percent of Total
SEX	Male	17	36
	Female	30	64
RACE	White	43	92
	Black	4	9
EDUCATION	High School	20	43
	Bachelor or AA	15	32
	Master's	9	19
	PhD	3	6
INCOME	0-5,000	5	11
	5,001-10,000	5	11
	10,001-25,000	15	32
	25,001-40,000	9	19
	40,001-75,000	10	21
	> 75,000	3	6
WHERE FROM	Gainesville	9	19
	Florida (not Gainesville)	14	30
	Outside Florida	21	45
	Missing	3	6
MARITAL STATUS	Married	27	61
	Not Married	17	39
	Missing	3	6
AGE	20-24	7	15
	25-34	22	47
	35-44	12	26
	45-54	3	6
	55-64	2	4
	65-70	1	2

Note Mean Age = 34 1 Standard Deviation = 10 6

Data Gathering on Informants and Choosing Alters

Before proceeding each informant was given an overview of the experiment and the types of information they would be asked to provide No informant refused to participate after hearing what would be required

The first module was designed to get information about the informant After providing information about themselves informants were asked to list the names of 60 people they knew (hereafter referred to as alters). A set of loose constraints was imposed on informants' choices to control for the varying definitions of knowing that each informant may have used. Following is a paraphrased version of what informants were told:

> You may use any person you know from anywhere or any time. Knowing in this experiment is mutual. That means that you know them and they know you, either by sight or by name. Your choices must be living and you must be able to contact them in some way, so they are not really active in your network.
>
> I prefer that you use real names--not that I will contact these people in any way, but in future modules you will be asked to provide information about each choice and it will be impossible for you to remember which imaginary names you associate with each person. In some cases though, you know people very well but cannot remember their first or last name, or you know no name at all. For example, you may know a grocery store clerk, but only on a first name basis. In these cases you can use a mnemonic name word which will make you recall this choice when you are presented the name again. Spelling is unimportant as long as you are consistent. The program will not take the same name twice so long as it is spelled in the same way each time.
>
> I prefer that you do not use children under the age of four as it is my experience that very young children have different conceptions of knowing than adults. Finally, I prefer that you use people who are more "active" in your network. That does not necessarily mean in frequency of contact or length of time you have known them. However, compared to a second grade teacher you have not talked to for 20 years, there are probably other choices that would be better.[12]
>
> When you get down to 20 choices remaining, stop for a moment to think if there is anyone obvious you are leaving out.

In providing these ground rules for the choices I tried to strike a balance between being too general or too specific. One of my primary objectives is to find categories or groupings that informants perceive in their global network. On the one hand it would be potentially contaminating to use too many examples or very detailed descriptions of knowing. On the other, it was essential to ensure that the informant understood his or her options. Potentially, each informant has a different definition of knowing, or more likely, many definitions. The narrative prompted people to use a very loose definition of their own. Some informants were frustrated, wanting more definition. This was consistently refused.

---

[12] In at least one case an informant interacted with her second grade teacher on a regular basis. This demonstrates how difficult it is to construct an example that will be neutral to everyone.

In light of my objective to discover perceptual subgroups which exist in a person's perceived global network, the question arises as to why only 60 alters were requested and whether these covered a wide enough range to effectively sample among the presumed subgroups. Sixty alters was the largest number any informant could tolerate, given the amount of information requested. In particular, the last module described below, increased in size quadratically based on the number of alters.

As to whether 60 is representative, that is a judgement the reader must make. Twenty six percent of the informants chose their alters in under 10 minutes, fifty eight percent between 10 and 20 minutes, and sixteen percent between 20 and 34 minutes. Several informants lingered on the last 10 to 20 choices.

In some cases informants did leave out very obvious choices. In one case, a married couple left each other out of their lists. This couple simply did not think of each other as an alter during the listing task. This is not a major concern in finding large groupings since it is assumed that others belonging to such groups will carry the weight in the analysis. That is, cliques will form based upon other members of the clique. However, depending on the alter, it could be critical in establishing the existence of smaller cliques and the relations between them.

The Autocorrelation Problem

Very early in the analysis of my data I became aware of the potential effects of autocorrelation. This comes from analyzing the alter data as individually independent data points, thus yielding an N of 2,820 rather than 47. Such analysis would be erroneous since each of 47 informants contributed 60 observations to this total of 2,820.

In some cases I did deal with the entire set as a whole, but this was limited to broad descriptions of the resulting sample. For instance, several tables describe demographic characteristics such as gender, race and educational background in percentages of the total of all alters. These are presented in conjunction with informant level analyses, and are there simply as a convenience for the reader. Questions such as "How many male alters were mentioned over all 47 informants?" seem obvious and should be addressed. The reader should interpret these tables with the potential for autocorrelation in mind.

The crucial analyses are all on the level of the informant. Observations of knowing theme types and cluster types were made with either with binary data, implying their presence or absence for an informant, or frequency data, implying the magnitude of their presence or absence for an informant. This avoids the problems stated above.

Ranking Alters on Levels of Knowing

Once informants had chosen 60 alters, they were asked to rate, on a scale of 1 to 5, how well they knew each of them. No matter what their definition of knowing may have been, a 1 signified the least of it and a 5 the most. If an informant knew an alter very well they used a 5.

Table 2 shows the usage of the five levels of knowing across all 2,820 alters. A "1" was the least used level, and "2" the second least used. This is to be expected since, given only 60 alters, informants are likely to choose those from their global network whom they know best.

Quite unexpected was the dip in usage of level 4. This might be an experiment effect, some informants having difficulty placing alters in some of the middle categories. Again, it is less likely for alters of low knowing levels to be chosen when there are only 60 of them, so level 3 may represent the level of entry for the majority of informants' alters.

Table 2 Distribution of knowing levels across all alters

Level of Knowing	Frequency	Percent
1	168	6
2	494	18
3	779	28
4	618	22
5	761	27

An examination of the average level of knowing for alters relative to their placement in the list of 60 reveals a distinct trend toward lower knowing levels as the informant goes through the list[13] Figure 1 is a graph of the average knowing level for alters across all informants So the left most point represents the average knowing level over all first choices The second point is the average knowing over all second choices, and so on A clear trend from higher levels of knowing to lower is evident This conclusion is further supported by an OLS procedure regressing placement of alters (1 to 60) in the original list against the average level of knowing for each place across all 47 informants An adjusted R-square of 68 suggests high explanatory power for the placement variable with a coefficient of 0.167 (significant at 0001)

To gain some insight into the variation in meanings of knowing between informants they were asked to define each of the five levels My instructions were to think about two or three examples of each level

---

[13] The ranking scale of 1 to 5 is clearly an ordinal scale Informant's may vary as to how they interpret the distance between 1 and 2 versus the distance between 4 and 5 Furthermore across informants these distances could also vary However the treatment of ordinal data as interval particularly those derived from scales is common practice in many of the human behavioral sciences For this research it is assumed that the values if 1 and 5 are standard across informants since they represent the extremes of the scale The variation in the intervals of values 2 3 and 4 is potentially problematic but the variation is more likely in how they are assigned Describing this variation in the assignment is one of the goals of the research

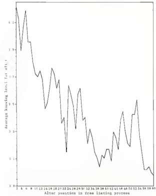

Figure 1 Plot of average knowing level by position of alter in free listing task

and the type of person that fall into that category The reader should refer back to the instructions given each informant concerning their choices It is possible that the mention of frequency of contact and duration of the relation might have biased informants towards these criteria

Table 3 lists the eleven themes found to be most prevalent in the 47 sets of descriptions On average informants conveyed 4 3 (SD 1 6) of these themes within their set of descriptions At first glance there appears to be a mixture Specifically, what are clearly groups such as "Family/Relative" contrast with attributes of knowing, such as "Frequency of contact " A somewhat vindicating observation for classic network research is the strong representation of the categories of friends family and acquaintance These three commonly used categories do seem to exist freely elicited in these informants' perceptions although the category friend remains analytically useless (Fischer 1982)

Table 3 Frequency/Percent of total that mentioned category in their description of knowing levels 1 to 5

Category	Frequency	Percent
Friendship	30	64
Frequency of contact	27	57
Know personal data about alter/confidant	26	55
Family/Relative	26	55
Closeness	24	51
Acquaintance	19	40
Duration of relationship	14	30
Know factual data about alter	14	30
Work/Business/School	13	28
Know through someone	6	13
Depth of discussions	4	9

Hammer (1984) cited differences between males and females in the listing of network choices based on duration of the relation. Specifically, males used less relations of long duration as network choices than females. Hammer concluded that the duration of a relation was not as important a criteria for males as females. Breaking the data in Table 3 by the sex of the informant, some discrepancies are evident. Most categories demonstrate equal representation between males and females. However a higher percentage of females mentioned friendship, family and duration of relation than males, while a higher percentage of males mentioned acquaintance and work as a theme in their usage of the 1 to 5 scale.

Further analysis of these frequencies appear in Table 4. Presented here are significant co-occurrences of themes, as measured by a Pearson's R, in order of significance. For instance, whether an alter is a friend tends to occur along with the theme of closeness, family and acquaintance. Examples of these are gradations from acquaintance to close friend or family. Another theme of knowing concerns the types of information the informant knows about an alter. So knowing factual data is more associated with lower levels of knowing, while knowing personal data is a higher knowing level theme. An unexpected result was the negative association of frequency of contact with the network "via." It is worth noting that there only six informants used the via concept in their descriptions of knowing levels.

A factor analysis on the frequencies of these themes yielded two factors which explained 37% of the variance in the 11 variables tested. The first factor loads the themes of family, closeness and friendship high, and the two knowing themes low using principal components. Rotating the factors either orthogonally or obliquely causes the lower loadings, which were not demonstrably strong in the original solution, to vanish. This leaves the themes of family, closeness and friendship defining the first factor which tends to be associated with high levels defining the first factor which tends to be associated with high levels of knowing

Table 4 Significant correlations (Pearson R) of categories

Cooccurrence of	With	Correlation	Prob>|R|
Friendship	Closeness	503	01
Friendship	Family/Relative	481	01
Family/Relative	Closeness	404	01
Friendship	Acquaintance	34	02
Know through someone	Frequency of contact	316	03
Know personal data about alter/confidant	Know factual data about alter	305	04
Duration of relationship	Frequency of contact	278	06

The second factor loads the via theme against frequency of contact and duration of relation This is sensible since the via relation is a consequence of another person Thus, frequency of contact and the duration of the via relation are also dependent on the intermediary alter The remaining factors accounted for little variance and loaded low

In Table 5 are the percent representation of a theme within a knowing level For example, 25.5% of the informants mentioned the theme of acquaintanceship in their description of their use of level 1. The category of 'Acquaintance' is clearly a low knowing level criterion as it is never used at levels 4 and 5 In contrast 'Family/Relative' occurs primarily at the highest levels, especially level 5 Clearly those who are family are not acquaintances

Less truistic is the use of the theme 'Frequency of contact' It appears at all levels with the exception of the low use at level 5 Frequency of contact is a theme which plays a part in all types of relations except the strongest Once a relation reaches this level, frequency of contact is of little importance Frequent contact is not necessary to maintain the relation 'Friendship Closeness' and

'Knowing personal data' tend to increase in usage as the knowing level increases

Table 5 Percent of knowing level containing theme

Category	K1	K2	K3	K4	K5
Acquaintance	26	15	4	0	0
Frequency of contact	23	34	36	21	6
Friendship	9	15	32	43	45
Know personal data about alter	9	11	13	36	57
Closeness	0	11	9	21	34
Family	2	0	13	15	45

Alter Detail

The next module was by far the longest taking some informants over two hours to complete For each of the 60 alters chosen the informant was presented with a screen prompting for several variables These are detailed in the section below Some non responses were allowed since it was unreasonable to expect informants to know all of the information about each alter Others were missing due to programming errors These are all represented by the category Missing

Not surprisingly the average age of all 2,820 alters was close to that of the informants (36 5 for alters SD 16 3 and 34 1 for informants SD 10 6) Over 75% of the alters fall between 18 and 54 and over 50% between 35 and 54 (see Table 6) Regressing informant age against age of alters yields a coefficient significant at 003 and an R square of 1732 This is not astounding as nearly 83% of the variance in alter age remains unexplained but it does show that informants tended to select alters close to their own age As was mentioned before, the fact that my sample is relatively young will yield biased results since informants tend to select alters close to their own age

Though 64% of the informants were women the distribution of males and females among alters was quite even (see Table 7) Males informant

Table 6 Age distribution of alters

Age	Frequency	Percent
0-5	115	4
6-17	82	3
18-24	324	12
25-34	895	32
35-44	624	22
45-54	389	14
55-64	176	6
65-74	154	6
75+	61	2

Note Average Age= 36.49 SD= 16.29

Table 7 Sex distribution of alters

Sex	Frequency	Percent
Male	1357	48
Female	1458	52

selected 56% male alters and 43% females while female informants selected 56% female alters and 43% males. A two-by-two crosstabulation of informant sex by alter sex results in a Chi square of 45.86 significant at .0001, suggesting a strong relationship. That is informants tend to select alters of their own gender.

To get a feel for the distribution of gender choices the data in Table 8 were computed. According to this table one informant listed 10-19 percent male alters, and 81-90 percent female alters. This was a female informant. Similarly, one informant listed 80-89 percent male alters and 11-20 percent female alters. This was a male informant. Referring to the table, the first two frequencies (1 and 3) are female informants while the last two (1 and 1) are male informants. The remaining categories are comprised of a mix of male and female informants. In general, informants' responses tend to approximate a

Table 8 Distribution of sex of alter across informants

Percent Alters Male	Female	Frequency	Percent
10-19	81-90	1	2
20-29	71-80	3	6
30-39	61-70	7	15
40-49	51-60	14	30
50-59	41-50	15	32
60-69	31-40	5	11
70-79	21-30	1	2
80-89	11-20	1	2

bell curve in their use of alter gender, with the majority of informants (61%) using equal numbers of male and female alters (±10%).

Alters, like informants, were overwhelmingly white (see Table 9). Of the 234 black alters, 181 (or 77.4% of all black alters) are accounted for by four black informants. This 181 (75.4%) represents a large portion of all of the 240 alters these four black informants listed. Black informants tended to list a higher proportion of whites (25%) than whites did blacks (1.9%). In general, informants selected alters within racial boundaries.

Table 9 Racial distribution of alters

Race	Frequency	Percent
White	2449	87
Black	234	8
Other	90	3
Missing	47	2

The distribution of educational levels is very similar between informants and alters (see Table 10 compared to Table 1) Most alters

Table 10 Educational distribution of alters

Education	Frequency	Percent
Elementary	119	4
High School	1172	42
Bachelor (includes RNs)	835	30
Master	411	15
PhD (includes MDs lawyers)	195	7
Missing	88	3

have a high school education as their highest educational level attained There was a problem with the coding scheme for education as it reflected a distinct bias on my part Many informants expressed frustration trying to respond for alters with two-year or vocational degrees Those still in school with no degree as yet were particularly problematic For all these cases, informants were instructed to code them as BAs For those seeking graduate degrees Master s degrees were used

The results of some non-standard questions are presented in Tables 11 through 17 In any social relation there are four possibilities regarding the strength of effort in maintaining it; A and B work equally hard A works harder than B B works harder than A, or neither A or B try to maintain it (see Table 11) In this set of social relations most are perceived by informants to be equally maintained by both, that is both informant and alter put about an equal amount of effort into keeping in touch The second largest category is situational knowing or ascribed relations, where neither informant or alter put effort into maintaining the contact The typical relationship of this type is the work place where two people are put into contact by virtue of duties which the work structure places on them although one informant used this coding for several family alters. Perceived asymmetric maintenance accounts for only 12.2% of the 2,820 informant/alter relations.

Table 11 Distribution of maintenance category

Who maintains relationship	Frequency	Percent
Informant	255	9
Alter	91	3
Both Equally	1835	65
Neither (situational)	591	21
Missing	48	2

Informants were asked to characterize their relation with the alter as positive, negative or neutral. A positive relation would suggest friendly or pleasurable interaction while a negative relation suggests an adversarial relation. Most of the relations were positive (77.4%) while very few were negative (2.8%), as seen in Table 12. However, several relations were perceived as neutral (17.2%)[14]

Table 12 Distribution of relational character

Character of relationship	Frequency	Percent
Positive	2181	77
Negative	80	3
Neutral	484	17
Missing	75	3

Informants were asked to rank how close, on a scale of 1 to 5, they felt to each alter, whatever that meant to them, completely independent of their ranking on the knowing scale. This question was added to shed some light on a debate as to whether knowing and closeness are the same thing If an informant is asked to list those who they are very close to will they also be alters who the informant knows well? A strong difference between knowing and closeness was not supported by the data A mean absolute difference of 63 (SD 80) a little over half a point demonstrates the similarity between the way knowing and closeness were used

---

[14] Note that these do not add to 100% due to missing responses

Table 13 is a crosstabulation of knowing by closeness This shows that 445 alters received a 5 on both knowing and closeness The bold values represent exact matches between the knowing and closeness rankings These account for 51% of all 2 820 alters In fact 90% of the alters were ranked ±1 in closeness from their knowing ranking, signaling a strong similarity in the way they were used (Note that closeness contains a 0 value. This was included since an informant might know an alter but not feel close to them This was the case 47 times)

Table 13 Crosstabulation of knowing rating by closeness rating

	Closeness						
Knowing	0	1	2	3	4	5	Total
1	8	96	54	7	3	0	168
2	12	97	**248**	126	11	0	494
3	10	51	207	**388**	**118**	5	779
4	8	12	47	236	**272**	43	618
5	9	15	27	65	**200**	**445**	761
Total	47	271	583	822	604	493	2820

Note Bold values represent exact matches on knowing and closeness

However of the 48 6% non-matching relations, 35 5% were of the form where knowing rated higher than closeness Intuitively this is explained by closeness being reserved for special people. You may know someone well but not feel close to them Again, the data suggest a minimal overall difference

Two questions from this section relate to findings from a series of experiments called the reverse small world (RSW) conducted by, Killworth, Bernard and McCarty (1984) and Bernard, Killworth and McCarty (1982). The RSW method is a variation of Milgram's 1970 small world experiment. In RSW, informants were presented with a list of 500 fictitious target people with seven pieces of information about each target. This information, such as where they live and their occupation, was stratified so they represented different geographic areas, and were balanced by gender, age, etcetera. Informants were asked to list the person they knew who would have the best chance of knowing the target, based on the information provided. By recording who informants used and the information about the target that cued them to make a choice, RSW suggested which aspects of a social relation were most pertinent in providing a connection to the informant.

From the RSW research, it became apparent that location and occupation were strongly related to the social relation as described in the RSW scenario. This led to the addition of two questions on this section to test whether the relations were location or occupation sensitive. The affirmative response to each question (e.g. 'Yes, I would know this person after a year') is suspect. People move with intentions of staying in touch, but fail to follow through. On the other hand, the negative response is probably valid as it is less likely for an informant to end up keeping in touch with someone they currently say they will not. The responses to these questions, listed in Tables 14 and 15, seem to replicate the RSW results in that 29.4% of the relations are location sensitive, and 16.2% occupation sensitive, thus suggesting that location is primary to the nature of their social relations. RSW (1984) resulted in 82% of "reasons for choosing a link" to be based on location.

These results agree with several studies which have demonstrated the importance of proximity in the formation of relations (Athanasiou

Table 14 Distribution of locational sensitivity of relation

If you moved over two days drive from Gainesville would you continue to know this person after a year?	Frequency	Percent
Yes	1990	71
No	830	29

Table 15 Distribution of occupational sensitivity of relation

If you changed occupations would you continue to know this person after a year?	Frequency	Percent
Yes	2363	84
No	457	16

and Yoshioka, 1973; Monge and Kirste, 1980; Sudman, 1988). We tend to know people with whom we come into physical contact. Although this seems obvious, there seems to be a common knowledge that innovations in communication technology significantly widen the scope of our personal networks. Studies such as these suggest that this effect is greatly exaggerated.

Table 16 and 17 show the results of asking informants where they met alters, and where alters live now, respectively. Most alters live in Gainesville and were met in Gainesville. Similar numbers were met and live in Florida and outside Florida. Crosstabulations of these two variables reveal that 84.3% of the 2,820 alters lived now and were met in the same location category. Thus most of them, to the informant's knowledge, are stationary.

Informants were also asked how they knew each alter. To avoid bias, the informant was first asked to give a short, textual response as

Table 16   Distribution of locations where informants met alters

Location where met	Frequency	Percent
Gainesville	1196	42
Florida (not Gainesville)	821	29
Outside Florida	751	27
Missing	52	2

Note Number of years known
Average=12 72 SD=16 03 Range=0 91

Table 17   Distribution of locations where alters are now located

Location now	Frequency	Percent
Gainesville	1134	40
Florida (not Gainesville)	813	29
Outside Florida	825	29
Missing	48	2

to how they knew their alters Since there was no way of checking each response as it was entered, and filtering responses presents its own biases the quality of the text was varied Informants were encouraged to type in what they wanted with the understanding that the information would be used to help us form groupings in the analysis They were discouraged from labelling someone simply as a friend

After informants provided the above information for each alter they were presented with the same list of 60 alters a third time To ensure a useful coding of how informants knew their alters, they were asked to provide at least one, and up to three, ways they knew each alter from a preset list of 23 ways people know each other (see Appendix) This list was compiled from a set of earlier experiments where informants were asked to describe how they knew their choices

This list was elaborated to some extent, providing slightly more detail for some categories which were originally too broad.

It is interesting to note that when allowed to free list, informants can fit most social relations into 23 categories, although some are admittedly broad. There were only a few instances where an informant felt compelled to use the 'Other' category, and most of these could easily have fit into an existing category. While the general relationships that were reflected in the textual descriptions were interpretable from the structured coding, the precoded categories frequently provided detail which was key for identifying clusters in the grouping analysis to be described in the next chapter.

Looking closer at the categories (see Table 16) there are several which could be grouped together, such as the relative relations or the work relations. It would be preferable, however, to have a set of relations that reflected a common theme, such as the function of the relation or special proximity. Many themes are reflected across the 23 categories listed, but no single one is dominant.

Two categories in particular deserve further explanation. The 'two step relation' occurs when the informant knows an alter through someone else. In everyday life, we frequently know people through other people. For instance, people may be introduced to a friend via another friend. But over time, the fact that they met through someone else becomes irrelevant. Were the via to become inactive, the step relation might still exist. A common example is two people who meet through a mutual friend and later become married. Their social relation with the mutual friend may become inactive, but their own social relation maintains. On the other hand, there are people who we meet through mutual friends where the via remains important. A typical example is in laws. A three step relation is a more uncommon situation where there are two intermediaries.

Table 18. Frequency of pre coded categories across all alters

Type of knowing	Frequency	Percent
Friend	1413	50
Two step relation	427	15
Co worker/colleague (not boss)	424	15
Acquaintance	422	15
Relative (blood, close)	389	14
Went/go to school together	366	13
Confidant	322	11
Relative (marriage, close)	186	7
Neighbor	185	7
Organization (same organization)	128	5
Three step relation	115	4
Hobby (involved in same hobby)	114	4
Boss/supervisor	113	4
Religious affiliation	100	4
Relative (marriage, distant)	74	3
Relative (blood, distant)	67	2
Service (supply some service for you)	65	2
Ex (spouse, lover, etc.)	64	2
Instructor/teacher	63	2
Romantically involved	30	1
Employee/supervisee	20	1
Student (your student)	12	0
Military	2	0

There are seven categories that are all used with at least 11% of the 2,820 alters. These are: friend, two step relation, coworker (not boss), acquaintance, relative (blood, close), school together and confidant. The presence of the high level of 'Went/go to school' is explained by the large number of university students who participated. The terms 'Friend' and 'Acquaintance' are attributes of a social relation that do not impart much information, although acquaintance suggests infrequent contact. Perhaps the most interesting piece of information from this table is the high number of two step relations informants listed. The 'via' as a form of social relation is apparently an important factor in the foundation of social structure. Keeping in mind that the informant could list up to three relation descriptors, it is also surprising that others, such as religious affiliation or organization, received little representation. Again, this may be a function of the high number of students in the sample.

Table 19 provides a breakdown of these categories sorted by average knowing level (1 to 5) and the distribution of the knowing levels associated with them. Thus, 90% of those alters who as relation was described as 'Romantically Involved' received a knowing level 5 from the informant and exhibited the highest average level of knowing (4.87), a result which one might expect. Acquaintances receive the lowest average knowing, although over 66% of them received a level 2 or 3. The category that received the highest percentage of knowing level 1 was instructor/teachers, although the existence of some strong relations between informants and teachers brings up the average.

Several categories display a more even distribution across knowing levels. These include neighbors, ex-spouse or lover, employees and friends. This demonstrates the variable character these relations can take compared to, say, a close blood relative. Close relatives by marriage rate higher than distant relatives by blood. Blood is not always thicker than water.

Pair Rankings

The final module of this experiment provided the data that formed the foundation for most of the analysis. Recalling that one goal of this research is to find informant defined groupings and attempt to explain them, this module created adjacency matrices for each informant which made such groupings possible. These matrices can be examined for subnetworks among informants' 60 alters.

Researchers have used a variety of methods in the past to create subgroupings from a list of elements. These methods tend to work at a higher level of abstraction than this method where the informant is presented with all the possible pairs in the 'global network' of 60 alters. Arguably, this abstraction might be desirable. If the task is to subdivide all 60 alters into cognitive groupings, the pile sort and

Table 19 Percent distribution of knowing levels for each pre coded category

Type of knowing	Mean	1	2	3	4	5
		Percent of Knowing Level				
Romantically involved	4.9	0	0	3	7	90
Relative (blood close)	4.6	1	2	6	14	76
Confidant	4.5	0	2	8	26	63
Relative (marriage close)	4.1	2	7	18	29	46
Friend	3.8	1	11	30	29	30
Ex (spouse lover etc)	3.7	5	13	30	16	38
Relative (blood distant)	3.6	6	8	40	18	28
Military	3.5	0	0	50	50	0
Religious affiliation	3.5	2	22	30	19	27
Hobby (involved in same hobby)	3.5	4	15	37	22	23
Went/go to school together	3.3	5	20	31	25	19
Relative (marriage distant)	3.3	10	14	39	18	20
Student (your student)	3.3	0	42	25	33	0
Employee/supervisee	3.2	10	25	20	25	20
Co worker/colleague (not boss)	3.2	6	20	40	21	14
Boss/supervisor	3.1	6	32	25	22	15
Two step relation	3.0	7	28	31	23	11
Three step relation	3.0	8	26	38	17	11
Neighbor	2.9	9	29	30	23	9
Instructor/teacher	2.9	24	14	29	13	14
Service (supply same service)	2.8	12	35	28	11	14
Organization (same organization)	2.7	18	25	34	16	7
Acquaintance	2.6	13	36	31	15	6

triad test tend to force the formation of these (though triad tests are impossible with more than 25 items since the instrument becomes prohibitively large) However in this case I wanted the informant's perception of the true network connection included in the formation of the grouping even if the perceived relations did not exist[15]

The method for setting up the task was to take an alter (X) from the list and present the informant with the question 'How well does X know Y' where Y is another alter on the list The scale was limited to 0 through 5 0 meaning they did not know each other and 1 through 5 being the same scale as was used by the informant to rank each alter

---

[15] Whether or not a social relation exists presents an interesting dilemma If an informant perceives a social relation and behaves contingent upon its existence in a sense it does exist Social scientists must think hard about operationalizing social relations only with independent objective proof of social contact

After the informant responded to each pair on the list using the first alter that alter was dropped and the second alter was compared to the remaining alters This continued until the informant had rated their perception of knowing between every possible pair

It has been established that the placement of alters within the list of 60 is significant (see Figure 1) A list of unique pairs constructed directly from this list in its original order would inherently contain the bias that high knowing people (that is people the informant knows well) would come first This could have affected the informants' responses on the first part overall Secondly most informants reported that they entered their 60 people in groups making the task easier Pairs created from the order in the original list would have created much more work for the informant in that proximate alters on the list would have a much higher likelihood of knowing each other Finally it is possible that informants grew weary of the procedure and got into a pattern in their responding This would potentially bias later pairings

A check on the responses does reveal differences The 1,770 responses for each informant were divided into halves and t tests run to see if the means were equal Overall the means were significantly (02) different with a mean of 67 (SD 1 40) for the first half and 7 (SD 1 43) for the second However, this includes values of zero It is reasonable to assume that the first half will contain more zeros than the second half since the first half is a more thorough coverage of each alter with all others Thus if an alter is a member of a small group or knows no other alters the effect in the first half will be more pronounced than in the second The same analysis excluding all zero entries yielded no significant difference, supporting this interpretation

In an effort to avoid the potential bias, real or imaginary the list of 60 alters was sorted prior to constructing the pairings The simplest method was to sort on the first names of actors since there was no reason to believe that this sorting order carries over into social relations. In other words, there is no reason to believe that Sarah knows Sam better than Albert just because of a difference in their first names.

Another problem with this module is with symmetry of the relations. Typically network analysts strive to obtain directed measures of social relations. That is, they allow for both symmetric and asymmetric ties. For example, observing the number of times two people communicate with each other is obviously symmetric. In many cases, however, knowing levels between two people are very likely asymmetric, at least in degree. That is, A may rank his knowing of B higher than B rates her knowing of A. This can cause problems in certain types of analysis. Asking how well one person knows the other may not be the same as the reverse question. Unfortunately, some of this could have been avoided by changing the wording, that is by saying, 'How well do these people know each other?', thus forcing symmetry.

Potential asymmetry is more likely a problem when informants rank how well they know each actor, since some actors may not agree on the knowing level assigned by the informant. But the pairwise ranking is actually a ranking of perceived knowing between two people, neither of whom is the informant. Asymmetry only becomes a problem if the informant happens to carry an asymmetric relationship between these people in his or her mind. Although it is possible for the informant to perceive asymmetric knowing among 1,770 pairings with a definition of knowing as broad as was used, it is unlikely to be a factor.[18]

Another unrelated criticism of this method is the tedium factor. When first described to colleagues, some said the task was simply too long and complicated to be reliable. What is more, this was after a data gathering module that took many informants two hours. In practice, however, the module went quickly, most people finishing in under 45 minutes, including error checks (errors that were written down and later corrected). As a reliability test, 30 more pairs were presented to the informant composed of every 59th pair from the original list. This was assumed to be random since the list was sorted by first name. In response to the criticism regarding asymmetry, the name order was reversed so the form of the question was "How well does Y know X?"

---

[18] Alba (1973) suggests that in certain cases assuming symmetric ties may be more realistic. This is particularly true in cases where informants can give only approximate answers, precisely the case in this experiment.

If there was a problem with asymmetry or if informants did tire of the procedure and became less accurate, this should show up in the additional 30 pairings as deviations from their original responses. As it turns out, informants were quite accurate.

A robust 88.7% of the rankings from the second set matched the first set. Adding in those where rankings were off by ±1, 96.5% are accounted for. Of course, most responses were 0 and it is much easier to reproduce a 0 than a scale between 1 and 5. Still, 65.6% of the previous non-zero responses are identical. Again adding in "mistakes" of one degree, 93% are accounted for.

The worst possible error would be claiming on the second round that two alters knew each other who did not on the first round, or vice versa. Out of 2,820 rankings, 78 (2.8%) were errors of this type. Forty five of these errors differed from the original ranking by one point, leaving 1.2% of the rankings differing from the original by more than one point. Since this experiment has not been done before, it is difficult to say if these figures imply accurate rankings or not.

The critical point is that some matrix procedures used in later chapters draw more heavily on the difference between 1 and 0 than 1 and 2, 2 and 4, etc. Fortunately, the procedure that was finally used for forming groupings relied on rankings of 4 and 5 which were quite accurate. Of the 1,350 checks, 27 were originally rated as a 4 or 5 and something different in the check task. Five of these were originally a 4 and later rated a 5, and another 5 were originally rated a five and later rated a 4. In any case, the final analysis would not have been affected. Since there is no way to know which ranking should be used (particularly if it is an artifact of asymmetry), the first round rankings will be used in all cases.

Network Dynamics

In the analysis of the data it became clear that the personal networks were quite volatile. Given my definition of knowing, it appeared that the content of many relations among informants' 60 alters was rather tenuous. At the risk of overstating the obvious, many relations were specific to location. Networks are also vulnerable to the entrance or exit of pivotal gatekeepers who mediate relations.

This is not to say that the 60 alters generated by informants be viewed as an ideal network, changes being considered deleterious. Yet it does point out how dynamic networks are, given some definition of knowing. They grow, shrink and change membership as time goes on, with some variable subset of constant relationships which can withstand the test of time, and changes in physical location or lifestyle.

As part of a verification process, after approximately six months ten informants were asked to rank the 60 alters they originally listed. Remembering that the accuracy of ranking knowing between alters was relatively high, the ranking of knowing of initial alters would have been even more accurate. Thus some differences between the original ranking and this one may be attributed to error, but the majority are more likely due to changes in the status between alter and informant. Comments by informants after the procedure verified this hypothesis. One informant, a dental student referring to a former patient, remarked that her knowing level was high while she was treating him, but fell dramatically after his treatment was complete.

Of the 600 re-rankings, 240 (40.0%) were exactly the same as the original ranking. A total of 483 (80.5%) were within 1 point of the original ranking, with the largest deviations being 3 points, both positive and negative. Overwhelmingly, the deviations suggest network attrition relative to the original list of 60 alters as 41.7% of the alters received a lower level of knowing than they did the first time, compared to 18.3% who received a higher ranking. The average knowing was significantly (.001) lower on the second occasion compared to the first. Assuming that the network is dynamic, it is likely that many of those who received lower rankings will continue to get lower, while others in the list will get higher. Still others will join the central core of "active" network members which is largely occupied by family[17].

Of the 240 alters who received the same ranking, 44.7% are relatives of the informants, while 28.9% of those whose rankings changed were relatives. This suggests that, for this set of informants, the core group is largely composed of relatives, as expected. The social relation between an informant and a relative is largely ascribed, and only secondarily contingent on other factors, such as location, work or common interests. The ascribed role of relative forms enough of a base to maintain a social relation over time and distance.

Summary of Alters

On average, informants' networks contained 52% females and 87% whites. The mean age was 37, with a standard deviation of 17 years. Regarding level of education, most alters had some college. They tended to be from Gainesville, but, in general, the informant met their alters where the alters now reside. Relations were overwhelmingly positive and most were equally maintained between the informant and the alter. Most relations were characterized as friendships, a category of questionable value

---

[17] Many informants suggested that family stand the test of time. Logically, family members are already likely to be long term relations for anyone over the age of 18. Thus to have been used in this experiment they are probably part of a stable set of social relations.

CHAPTER 5
GROUPING ANALYSIS

Clustering Procedure

By assuming symmetric measures of knowing from the results of the previous section, it was possible to construct adjacency matrices for each of the 47 informants. Including their own initial assessment of how well they knew each of their 60 alters, each matrix ended up with 61 rows and 61 columns. Given the assumption of symmetry, the lower and upper diagonal halves of the matrix were identical, consisting of values ranging from 0 through 5, representing how well two alters knew each other. All values along the diagonal were fives, reflecting the fact that each informant knew themselves to the maximum.[18] This matrix is easily input into a variety of analytical procedures, some of which are network specific.

One of the first procedures used to look for subgroupings in my data was cluster analysis, a technique for separating data into groups whose members are somehow more alike than they are with members of other groups. A thorough treatment of cluster analysis is found in Everitt (1974) or Jain and Dubes (1988).

As with many complex statistical procedures, there are many types of cluster analyses. One such analysis, the ADCLUS model of Arabie et al., was chosen because it allows for overlapping clusters, that is, alters can be members of more than one cluster. ADCLUS produces a series of groups with an overall R^2 for the fit and estimated explained variance for each group. As with all exploratory analysis interpretation plays a large part in deciding if groups are sensible or not. In this case, the amount of variance explained must be balanced with the number of members in each group to determine if the group is significant. For instance, a group of three which accounts for 3% of the overall variance may be just as significant as a group of 20 that accounts for 10%.

---

[18] The diagonal values are unimportant in many analyses since they are excluded. For those procedures that include diagonal values, the question arises as to whether they should be higher than any other possible value, since informants most likely know themselves more than anyone else. Cluster analysis assumes symmetry and thus ignores both the diagonal values and one diagonal half of the proximity matrix.

The data associated with each choice were used to make sense of the groups. Most important were the textual description of each alter, the structured knowing ranking, and in some cases the last names. It is entirely possible that significant groups have been left out due to subtle relationships which were not apparent given the information elicited about each choice. Less likely is the identification of a grouping that the informant does not see as associated. A test for this type of error was given to a subset of informants, the results appearing later in this chapter.

Finally, the reader should always bear in mind the nature of the adjacency matrix. Reflected in the data are groupings whose members are more likely to know each other than they are to know people in other groups. This is true no matter how the informant interpreted "knowing." What may not show up are categories where groups are created by attributes rather than by actual contact. Whereas a pile sort or triad test under the instruction to sort by similarity alone might yield a group of 'girlfriends' or 'boyfriends', this is not a group that would be expected to know each other, and therefore is not likely to appear in my results.

Most of the interpretation of clusters is based on the textual description of each alter provided by the informant. Since it was impossible to screen these data by computer, and it was potentially biased to direct informants on the content of these descriptions, some were better than others. In many cases I used a combination of the structured knowing categories and the textual descriptions to determine why alters clustered together.

Two informants provided such sketchy descriptions and non varying structured categories that it was impossible to decipher why any of their clusters existed[18]. Three other informants provided descriptions that were difficult to interpret, but some clusters were sensible. It is very likely that other informants' descriptions may also have been lacking the detail needed to make sense of a possibly sensible cluster. Thus the number of clusters probably errs on the conservative side. That is, there are probably more sensible clusters than are reflected in my results since care was taken not to select clusters where a very reasonable mechanism for members knowing one another was not apparent. Since two informants provided non varying detail, they were excluded from the summary below, leaving 45 usable matrices.

Cluster Types

In coding the 45 sets of clusters, 12 major categories were distinguished, three of which were further divided, for a total of 19 cluster types. Table 20 shows the types and the percentage of informants who had clusters corresponding to these types. There are actually two sets of percentages displayed reflecting two approaches to cluster definition. The difference between these will be discussed in the following section.

Some very common types, such as family, work and school, are strongly represented in the list of types. Perhaps less truistic are the high frequency of couples, social clusters, friend/family clusters and significant other's family. Upon first examination these seem

---

[18] One of these was a male, age 29, whose alter descriptions were nearly all "friend met in night club". In all likelihood this was true. Given more detail it may have been possible to subdivide these alters further, perhaps by different nightclubs. However, the descriptions provided made this impossible.

Table 20 Percentage of informants who exhibited each cluster type for each analysis

Cluster Category	No ego/binary Frequency	No ego/binary Percent	Ego/raw Frequency	Ego/raw Percent
Family	45	100	43	96
Maternal	6	13	7	16
Paternal	7	16	6	13
Close	32	71	23	51
General	28	62	27	60
Significant other's family	22	49	15	33
Including friends	25	56	27	60
Network via other person	25	56	21	47
Couples	23	51	13	29
Childhood, growing up	4	9	2	4
School together	21	47	19	42
High School	5	11	8	18
University	18	40	16	36
Work together	33	73	33	73
Current work	23	51	22	49
Former work	19	42	18	40
Housemates their acquaintances	3	7	3	7
Religious affiliation	5	11	4	9
Hobby group	6	13	5	11
Issue oriented group	8	18	6	13
Neighbors	13	29	10	22
Social group	13	29	8	18

obvious. Indeed, if they were extremely obscure their validity would be suspect. But the latter four mentioned are not normally represented in network experiments.

The subdivision of school into high school and college is clear as is work into former versus current jobs. School, like work, is a more general type of social relation than the more specific subcategories which reflect differences in membership according to time period and location. The family clusters demonstrate some detail that is not expected. In particular, family including friends of family was a distinct type that over half of the informants produced. The decision to categorize the cluster as family including friends was reserved for those clusters where family represented a strong majority of the members along with others who were described as family friends. Thinking back to my own childhood, certain friends were an important part of our family affairs These were adult friends of my parents who themselves had children with whom my family interacted frequently

Some informants perceive their family to be divided into maternal and paternal clusters However in most cases they did not occur together I attribute this to the mobility which typifies people and families today Given job constraints many people find it difficult to maintain relations with both parents' relatives It is even more difficult for married people or those with significant others[20], who are expected to maintain relations as a unit yet actually double their potential responsibilities Assuming limited resources, mainly time, some relations will most likely deteriorate over time due to low levels of maintenance. As to which relations deteriorate is probably affected most by distance if my paternal aunt lives in the same town as I do we may develop a stronger relation than would exist with my sister who lives in another state

Another subdivision of types was family based on the feeling of distance in most cases this was similar to levels of knowing or closeness Close family was reserved for clusters where nearly all members were defined as close family General family included both close and distant family These cluster types occurred proportionately more than any other Like friendship, they describe an attribute of the relations of the members to the informant, above the social relation of family

Networks via other people and couples are closely linked types Couples were counted as a clique since my assumption was that they would ultimately include the informant creating a triad[21] Most network vias are composed of a couple and their family The type includes clusters where the informant knows only one member strongly This is frequently the case with couples as is demonstrated by the high number of alters (194) who are two or three step relations away from the informant

---

[20] Assuming that this extra burden is a phenomena of marriage would be erroneous Crosstabs of marital status by presence of clusters composed of significant other's family reveal an even distribution across cells and an insignificant Chi square

Housemates and their acquaintances are also similar to network vias if not a special case of a via They are distinguished because the social relation which draws in the informant is a more specific type than the general type presented by the network via Indeed in some sense significant other s family might also be a via since a severing of the significant other relation more often than not would result in the severing of social relations with most or all of the cluster members

Finally religious hobby social and neighbor clusters are self explanatory Examples of issue oriented clusters are therapy or encounter groups and political or group rights organizations The reader should realize by now that these are not mutually exclusive types Members can and are members of more than one cluster type The process of developing these type was inductive and subjective based on the predominant set of relations within the cluster My interpretations were largely corroborated by the rest of clusters to be described below

On average informants expressed 6 33 (SD 1 82) cluster types in the 14 possible clusters A minimum of 3 and a maximum of 11 cluster types were used The maximum of 11 cluster types came from my own case This is not related to my being different from many other informants More likely it is related to my knowledge of myself and my ability to interpret my clusters with greater validity. What is more, I probably picked my 60 alters in an attempt to cross many types of people in my network, thus generating more groups. Another informant evinced ten cluster types. This was a former coworker with whom I continue to have contact. Again, I know more about her life than the lives of most of the informants. Three informants used only three cluster types, and all three informants were very stingy in their alter descriptions.

---

[21] All network definitions of clique assume a minimum of three members This seems odd since networks are composed of dyads Excluding a highly communicative and perceptually distinct link from clique definition is too restrictive

To conclude, the cluster types, and particularly the number of cluster types, is very much related to the detail of the alter descriptions. More might be gained by interviewing informants about alters in an unstructured manner where a minimum amount of detail would be assured. Of course this would require more time per informant and assistance with the research.

Age/raw Versus Non-age/binary

As was mentioned above, Table 20 contains two sets of frequencies corresponding to two different sets of criteria for defining a cluster. The 'age/raw' data sets were derived from the raw data matrices including all values (i.e., 0 through 5), and with a row and column representing how well the informant knew each alter. The 'non-age/binary' data set is closer to the Luce-Perry (1949) and n-clique (Luce, 1950) models in that the data were converted to binary and age was excluded. All values of 1 to 3 were recoded as 0 and values of 4 and 5 were recoded as 1. As can be seen from the table, the frequency of occurrence of these types is similar no matter which method was used.

Table 21 presents a further analysis of the differences. The entries for 'All differences' represent the total number of differences for that category between the two analyses. For example, there were a total of 12 cases where couples was an existing cluster type for an informant in one analysis, but not the other. Entries for 'non-age/binary' represent cases where a cluster type existed in the non-age/binary analysis, but not in the age/raw. In turn, the age/raw

Table 21 Frequency of mismatches of cluster type presence between data matrices

	All differences	In non age binary	In age raw
Family Maternal	0	0	0
Family, Paternal	1	1	0
Family, Close	3	3	0
Family, General	4	1	3
Significant other's family	6	6	0
Family, including friends	11	5	6
Network via other person	5	4	1
Couples	12	10	2
Childhood growing up	2	2	0
High School	1	0	1
University	1	1	0
Current work	3	2	1
Former work	3	2	1
Housemates their acquaintances	2	1	1
Religious affiliation	1	1	0
Hobby group	1	1	0
Issue oriented group	0	0	0
Neighbors	3	3	0
Social group	4	3	1
Total differences	63	46	17
Percent of all knowings	7	5	2

entries are those cluster types present in that analysis for an informant but not in non age/binary for the same informant

Of the 855 possible matches there were only 63 differences that is 7 37% of the total Although this does not seem radically different the reader should note that these are differences in the presence of a category rather than a frequency of specific occasions of differences

The fact that there are differences is explained by the noise which the 1 through 3 values present in the age/raw analysis These cells must be taken into account by the ADCLUS model and low level relations are collapsed into otherwise strong clusters The non-age/binary analysis has the effect of having forced the informant to truly think of who had a strong knowing relationship Contrasting 1 through 3 with 4 and 5, and then recoding them eliminates low level knowings that would still be coded as a 1 if the informant were asked to indicate knowing on a binary scale. Eliminating ego should have the effect of removing those alters who are tied to a cluster only by their relation with ego and no other alter.

Most of the differences are accounted for by cluster types present in the non ego/binary data and not in the ego/raw data, a total of 46 mismatches. Of these, most appear in the couples categories as well as family categories. This is sensible since couples and families are likely to be perceived as very closely knit, while hobby groups would not. So a hobby category is likely to exist in either analysis as it is a low knowing level cluster, while couple clusters might not appear unless the high knowing levels are exaggerated, given their size with respect to other cluster types.

On the other hand, there are categories which may exist at a low level of knowing but not a higher one. For instance, Family general, which means a mix of close and distant relatives, does account for three mismatches. The cluster type "Family including friends" is split almost evenly between the two analyses. This may be because some family friends are very close to family, and others are less close. These mismatches present in the ego/raw analysis tend to be associated with general family, indicating the presence of distant relatives. The mismatches present in the non-ego/binary analysis tend to be associated with close family.

On average, 23% of informants' possible clusters did not contain ego. In other words, the inter knowing between those alters was strong enough that the presence of the informant was not necessary for their formation. Frequencies of the types of these non ego clusters reveals little pattern. Most cluster types are represented in the non-ego clusters. There is a slight tendency towards close and general family, 24 of the 71 non ego clusters falling in this category. Yet this is hardly significant.

Since the two types of analyses, raw data including age versus binary data excluding age, represent radically different approaches, yet produce similar results, the remaining analysis will be restricted to one data set. Of the two, the non age/binary generated more clusters and yields slightly more interesting results. Most of the following analysis will refer to these data.

Verification of Clusters

A common criticism of many multivariate techniques is that the results will always generate something sensible. In this case, some might say that any cluster of alters can, in some way, be shown to be a sensible group. This is due to the exploratory nature of the analysis and the absence of theory driving the cluster formation. To test this proposition, a subset of ten informants were presented with the groupings described above along with a set of random clusters. Thus each informant was presented with a card which consisted of a cluster formed by the ADCLUS model, or one of seven clusters generated by selecting random numbers between 1 and 60 to fill clusters of 2, 4, 6, 8, 10, 12 and 14 alters.

Every cluster formed by the ADCLUS model was recognized by the respective informant as an interacting group, and a reason for the inter-knowing was given which was in agreement with my interpretation. There were only 6 randomly formed groups out of 70 which were identified as meaningful clusters, an average of .6 for each informant. Two of these six were comprised of two alters. It is much easier to pick randomly two alters who knew each other from a free list of 60 people than some larger group. Nevertheless, these were identified as sensible. Three of the six were interpreted as meaningful with qualifications. In other words, they were meaningful if one or two alters were excluded. The last of these six random clusters which were interpreted as meaningful was generated by an informant with a very dense network who used mostly family and neighbors. Again, a random grab of alters from a list of alters where the type of social relation is the same and relations are dense is more likely to yield a meaningful cluster than a random grab from a list which exhibits more variability in the social relations and less density.

It is safe to conclude that the ADCLUS model generated sensible clusters, and that my interpretation of these agrees with the perceptions of the informants. Furthermore, random clusterings of alters are overwhelmingly not sensible.

Variance Explained

Each informant was limited to a total of 14 clusters in the ADCLUS analysis, creating the possibility of 630 clusters for the 45 informants whose data were useful. Of 630 possible clusters, 449 (71.27%) were deemed meaningful. Of these 449 clusters, the minimum membership was 2, the maximum was 25 and the average was 5.35 (SD 3.7). Each cluster had an estimated proportion of variance explained[28]; the minimum was .27%, the maximum 69.83%, and the average 6.26% (SD 9.39%) per cluster.

Forty clusters (8.9% of the 449 meaningful clusters) explained more than 50% of informants' variance. That is, they were key clusters for their respective informants. Most informants (89%) had such a cluster.

These strong clusters were entirely accounted for by the cluster types of close family (n=17, 42.5%) and university (n=23, 57.5%). This suggests that when these categories were present, they were very important and central to the informant, relative to other cluster types. One example is the cluster of a student in dental school. She had a university cluster that explained 69.83% of the total variance for her

---

[28] The SAS version of the ADCLUS model produces a statistic, called the "Type II proportion of variance explained", which is the Type II sum of squares divided by the total sum of squares. The Type II sum of squares are the reduction in error sum of squares due to adding a cluster after all other clusters have been added. This serves to treat each cluster independently in evaluating the reduction of error, given the potential of overlapping clusters.

14 clusters Not surprisingly, this cluster was comprised of dental students Dental school like other professional health programs demands nearly all of a student's time for three years This makes the dental class virtually the only vehicle for socializing during this time

Moving to the informant level there were an average of 9 98 (SD 2 83) meaningful clusters per informant with a minimum of 2 and (per force) a maximum of 14 Totalling the variance explained for meaningful clusters an average of 62 5% (SD 18 38%) was accounted for per informant by their meaningful clusters with a minimum of 13.39% and a maximum of 97 57%

As expected cluster size is positively and significantly related to the amount of variance explained by a cluster--the more members in a cluster the more variance is explained

Informant Characteristics and Clusters

A reasonable question to ask is whether the attributes of informants in any way explains the attributes of the clusters Caution must be taken in doing this since the sample of informants is skewed towards 20 to 30 year olds many who are students All caveats aside this analysis demonstrated little explanatory power for informant characteristics

There are several potential dependent variables to test These are number of clusters average cluster size total variance explained and the presence or absence of the 19 cluster types Many of the informant characteristics appear to offer potential explanatory power for these dependent measures These are gender age race, levels of education and income marital status and if the informant was originally from Florida or elsewhere Given the small number of informants t-tests were used to test whether differences existed between binary explanatory groups (such as gender or race) on continuous variables, while correlations, OLS and logistic regressions were more appropriate for testing the explanatory power of binary dependent variables

Most informant characteristics were useless in explaining the continuous variance in number of clusters, average size and total variance explained. Only race specifically whether the informant was black or not was useful in predicting these variables If the informant was black total variance explained was significantly (.01) lower while average cluster size was significantly (.01) higher Recalling that there were only four black informants the strength of this result is questionable

Explaining the presence or absence of a type of cluster by characteristics of the informant was largely uninteresting Although Table 20 shows variance in the distribution of cluster types, the explanation for their formation does not appear to be a demographic one relative to the informant At least for these informants, gender race and other socioeconomic attributes do not do a good job of explaining the presence of cluster types There were exceptions, i.e., significant variables but given the number of tests it is questionable whether these are truly meaningful or due to chance

Table 22 details the notable associations between informant socio demographic characteristics and the presence of certain cluster types The category Know type 2' refers to an estimate the informant made as to the number of people they could call on if they needed help such as with child care or money matters This has been referred to as the instrumental network[23]

Most of these associations make intuitive sense For example, the higher an informant's education level, the more likely he or she is to have a college type cluster Less intuitive are the correlations

---

[23] Actually, referring to this as a network runs counter to the spirit of this research There is no reason to believe that the set of alters an informant can rely on for these things is a distinct, interacting subgroup

Table 22 Associations between informant characteristics and the presence of cluster types

Informant Variable	Cluster Type	Association	Prob>R /ASE
Age	Religious	0 52(R)	01
Education level	College	0 39(D)	09
Know type 2	Family including friends	0 50(R)	02
Know type 2	Maternal family	0 38(R)	01
Age	Family including friends	0 37(R)	.01
Education level	Hobby	0 62(G)	24

Note Depending on the level of measurement a Pearson's R Somer's D or Gamma were used as measures of association For Somer's D and Gamma the ASE (asymptotic standard error) is provided

involving informant age Older informants were more likely to have religious clusters but less likely to have clusters of family which include friends of family Conceptually, friends of the family is a perception of the child in relation to adult family members If a couple is friends with my parents, the American kin system does not offer a special name for them To distinguish them from others they become friends of the family To my parents they are friends, neighbors or some other cluster type To me they are friends of the family Thus the older the informant, the further away they are from social relations that center around the family and the family friend social relation

The type 2 knowing correlations present contrasting results The larger the informant's estimate of the people he or she can call on for help the more likely they are to have maternal family clusters although they are less likely to have family clusters including friends Maternal clusters are probably sources for people who could be called on for help, explaining the maternal correlation Yet the negative relationship with family friends is curious

Within Cluster Similarity

Recall that a large part of the experiment involved collecting dossiers on each site At the time of the instrument design, it was assumed these data would be useful for explaining why clusters formed as they did Would the age of cluster members or their educational level be useful in explaining their high level of inter knowing? Alternatively, are members of a cluster more similar on selected demographic variables than those outside the cluster?

In order to test this the data were reduced such that each alter had 14 dummy variables indicating membership, or not, within any of the 14 potential clusters per informant The SAS GLM procedure was repeated for each informant to explain variation in their alters ages by these 14 variables This constituted an unbalanced ANOVA given the simplicity of the models Note that informants varied on the number of dummy variables used since not all clusters were ultimately determined meaningful

From this procedure the R squares were averaged across informants This represents an average proportionate reduction in error per informant due to knowledge of group membership In addition to the R-squares each group coefficient had a T value associated and a significance value for that T These group coefficients with significance values less than 05 were presumed to have contributed significantly to the model

On average each informant s model had an R-square of 34 (SD 16), with 1 8 (SD 1 3) groups contributing to the model This is by no means a large value, although the R squares ranged from 01 to 73 The conclusion must be that over all groups ages are not significantly similar although within some groups they clearly are

Of the 77 groups that contributed to the model in explaining alter age 65% were some type of family cluster Another 10% were network via clusters which are frequently family The conclusion is that many clusters that have close ages relative to outsiders are family clusters Observing that family members tend to be close in age is not an interesting finding Overall it seems unlikely that alter age would serve to explain cluster membership over the type of social relation in this case family.

Given the utility of alter age in this analysis, it seemed futile to use other variables to repeat this test. Since most of the remaining dyadic variables are nominal or ordinal, they lack the explanatory power of ratio variables, such as age. If age performed so poorly, it is unlikely that these other variables would perform better.

Correlations

More interesting are the correlations between categories summarized in Table 23. Results for analyses on a matrix of frequency data versus binary data are presented. The binary analysis represents the presence or absence of cluster types for each informant, while the frequency data represents the number of times a cluster type appeared for each informant.

Oddly, there was only one common correlation between the two analyses. Thus when weighted for the number of times a cluster type occurs, the associations between clusters changes dramatically. This suggests significant variability in the frequency of the cooccurrence of cluster types, a variability which is removed when looking only at their presence or absence (binary values). This might also explain why frequency data generates less correlations.

A review of this table shows that some of these correlations are more easily explained than others. The first, for instance, is understandable in the sense that those who associate with a significant other's family, most likely a spouse, have more close family to form close family clusters. A negative correlation between close and general family suggests that there is a tendency for those with close family not to have general family, rather than having both close and general family, which would be a two tiered structure. Could this imply that some informants feel that all family are close, categorically?

Table 23 Correlation of cluster types

BINARY VALUES

Cluster Type 1	Cluster Type 2	R	Prob>R
Close family	Significant other's family	43	01
Former job	Family, including friends	- 32	03
Hobby group	Issue oriented group	33	03
Hobby group	Close family	- 33	03
General family	Network via other person	33	03
College	Neighbor	32	03
Hobby group	General family	31	04
Housemates	Former job	31	04
General family	Religious group	31	04
High school	Former job	30	04
Issue oriented group	Network via other person	30	05
Close family	General family	- 29	05
College	Religious group	29	05
College	Current work	29	05
Housemates	Network via other person	- 30	05

FREQUENCY VALUES

Cluster Type 1	Cluster Type 2	R	Prob>R
High school	Maternal family	36	01
High school	Family, including friends	43	01
Childhood friends	Housemates	40	01
High school	Network via other person	35	02
Family, inc friends	Network via other person	32	03
Hobby group	Social group	33	03
College	Social group	31	04
Paternal family	General family	29	05

The negative correlation between close family and hobby groups is confusing, unless it suggests that those who participate in hobby groups seek these as a surrogate for close family This interpretation is supported by the positive correlation between hobby and social groups suggesting that these hobby groups are places where informants socialize perhaps as they might with close family

College groups are negatively associated with neighbors religious groups and current work College students in general are less likely to have jobs, given their studies or to be heavily involved in church groups and given their transitive living conditions less likely to stay in the same living unit for more than a year and thus forming groups comprised of neighbors They are however very likely to socialize with other college students yet the positive correlation with social groups from the frequency data.

High school groups tend to cooccur with maternal family, as well as family including family friends and networks via other people. This creates a picture of an open family where the formation of subnetworks at a high school age is easy. There were only two cases out of 45 where maternal family was coincident with paternal family. In general, it seems that informants' families tend to gravitate towards one side of the family or the other. The positive relation between the presence of paternal family and a mix of close and distant family could mean that paternally oriented families are less open than maternal families. The negative association between high school groups and former job is clearly related to time. Those who are close enough to high school that they maintain these groups are less likely to have had the experience of having a former job, or at least one of any social consequence.

Factor Analysis on Cluster Type

For both the ego/raw and non-ego/binary data sets, an incidence matrix was constructed where a "1" represented the presence of a cluster type for an informant. This 19x45 matrix was submitted to factor analysis and multi dimensional scaling procedures. The MDS yielded no discernible groupings of cluster types despite subsequent analyses using the frequency of cluster types as well as the relative weights for each cluster generated by ADCLUS. The factor analysis yielded mixed results.

Four factors accounted for 46% of the total variance, not a startling result. However, the composition of some factors was interesting. The best solution was obtained using oblique rotation. The first factor, accounting for 14% of the variance, loaded close family and significant other's family against general family and hobby clusters. Note that these loadings range from .6 to .7, which might not be considered strong enough given the low variance explained. Again, the concept of high knowing versus low knowing clusters is consistently demonstrated A t test between the high and low loading clusters as groups suggests that the high loading clusters have significantly (02) higher average knowing (4 34) than the low loading clusters (4 19)

The second factor accounting for 12% of the variance is more subtle This loads paternal family and former work clusters high against family including friends A straight principal components analysis also loads maternal family high although this disappears using any method of rotation Maternal family's disappearance is very likely due to its strong association with high knowing clusters (i e the first factor) Excluding the former work clusters this factor has an obvious interpretation Family including friends is exclusively an immediate family cluster That is, family friends are always associated with an informant s parents and siblings The contrast with paternal family suggests a dimension of family clusters opposing immediate family with non-immediate relatives Unfortunately the inclusion of former work which loads nearly as high as paternal family in this factor is baffling

Factor 3 accounting for 11% of the variance, loads college high school childhood and housemate clusters against current work This is most likely reflecting the negative correlations between students and work That is the high loading clusters tend to be associated with students Many students do not work making it impossible to form work clusters Those who have work tend to do so part time often in retail or service jobs where it is difficult to form work clusters which depend on day to day interaction

Similarly the fourth factor accounting for 9% of the variance loads college clusters against neighbor and religion clusters Again this is sensible given the tendency of college students to be less associated with religious groups than older working informants as well as their living situations which tend to be temporary making formation of neighbor clusters difficult The negative relation between college clusters and work religious and neighbor clusters is reflected in the correlation analysis above

Unlike the factors on knowing themes these results are arguably weak On the one hand the explanation of 46% of the total variance is not high But given the type of variance being explained, that is the presence or absence of cluster types this should be considered notable The ambiguity of the results is directly attributable to the gray scale between certain cluster categories which in turn derives from the scant descriptions some informants gave of their alter relations

Overlap

The main difference in using the ADCLUS model over other clustering procedures is the flexibility to have alters fall in more than one cluster Although hierarchical procedures do allow alters to fall in more than one cluster at various levels a primary division early in the procedure will prevent alters from moving to another set of clusters For example if a hierarchical procedure initially divides alters into family and non-family it would preclude a family alter's membership in a cluster that is derived from a non family base In essence, the overlapping method is only an advantage if it results in unexpected category overlap

Table 24 shows the frequency of incidence of cluster types for each informant where a cluster type is counted if an informant has an overlapping alter who falls in that category So sister X, who falls into a close family cluster and a general family cluster would alone generate the incidence of these two types for this analysis This table does not reflect cases where an alter overlapped into another cluster of the same type So an alter who appears only in two close family clusters would not generate the incidence of close family in Table 24 Such overlaps were excluded as they were potential subsets of existing clusters A total of 39 of the 45 informants listed alters who overlapped between different cluster types

Table 24 Frequency of incidence of cluster types that contain overlapping alters for an informant

Cluster Type	Frequency	Percent
Family, Close	28	72
Family General	23	59
Family including friends	21	54
Network via other person	19	49
Significant other's family	15	39
Couples	12	31
Social group	11	28
Current work	10	26
Former work	9	23
University	8	21
Family Maternal	6	15
Family Paternal	6	15
Neighbors	6	15
Religious affiliation	4	10
Childhood, growing up	2	5
Hobby group	2	5
High School	1	3
Housemates their acquaintances	1	3
Issue oriented group	0	0

Over the 39 informants who listed alters who overlapped across cluster types there were on average 6 36 (SD 4 65) overlapping alters per informant Overlapping alters do exhibit some characteristics which distinguish them from non-overlapping alters They are significantly (003) older with an average age of 39 5 versus 36 2 and have known the informant for significantly (001) longer than non overlapping alters (19 6 compared to 12 1) Overlapping alters rank significantly (.001) higher on both the knowing and closeness scales than non-overlapping alters, 4 5 to 3 4 for knowing and 4 3 to 3 0 for closeness

Table 24 makes clear the dominance of the family categories for this type of alter. In other words alters who tend to exist in more than one cluster type, also tend to exist in some type of family cluster Table 25 takes this analysis one step further This shows the highest frequencies of cluster type pairs resulting from overlapping alters For instance there were 16 informants who had alters that were clustered in close family as well as general family Again it is difficult not to notice the dominance of family categories in this table

Table 25 Frequency of most common combinations of overlapping cluster types where overlap is across different types

Category 1	Category 2	Frequency
Family Close	Family General	16
Family, Close	Significant other's family	15
Family Close	Family, including friends	15
Family Close	Network via other person	15
Family General	Family including friends	12
Family Close	Couples	11
Family inc friends	Network via other person	11
Family General	Significant other's family	8
Family Close	Former work	8
Family General	Network via other person	8
Family General	Social group	8
Significant other's family	Network via other person	8

These tables may be interpreted in two ways The first and more uninteresting is to question the cluster types used for coding due to the high number of overlaps attributed to family clusters Since these were subjective types this could suggest that the categories are not as independent as first thought if alters freely overlap between close and general family, for example, then perhaps the categories are not different There are two arguments against this interpretation Again looking at Tables 24 and 25, the frequencies in general are not terribly high If the categories were not distinct the expected frequency of overlap would be higher Second are the significant negative correlations between many of the family type categories, suggesting that they are indeed different

Another interpretation which borders on the truistic is that family members exist in many cluster types because family categories dominate the list of 19 types Thus there is more opportunity of overlap between family categories They still represent distinct types but are highly integrated Of the two interpretations, this appears the most viable. If the free-listing task was constrained to non-family alters, the frequencies displayed in Tables 24 and 25 might be more interesting and less obvious.

Table 26 is similar to 25 except that at least one cluster type is not a family cluster. With the exception of two cases of via networks

Table 26. Frequency of most common combinations of overlapping cluster types, where overlap is across different types, at least one of which is not a family cluster

Category 1	Category 2	Frequency
Family, Close	Couples	11
Family, Close	Former work	8
Family, General	Social group	8
Network via other person	Couples	7
Significant other's family	Couples	7
Family, General	Current work	7
Family, General	Couple	7
Network via other person	Social group	6
Family, including friends	Current work	6
Family, including friends	Couples	6
Significant other's family	Former work	6
Family, General	Former work	6
Family, Close	Current work	6

all combinations consist of at least one family cluster. However, the cooccurring type is almost completely accounted for by couples or some form of work cluster. This suggests that some informants' family members have also interacted, or at least know, one or more members of couples and work groups the informant has maintained. Recall that this is not simple knowing, but enough knowing for ADCLUS to have clustered them together at a high level of knowing. So, the eight combinations of close family clusters and former work clusters suggest common members who exhibit high knowing levels in both groups.

Table 26 was arbitrarily cut off at frequencies of six. That is, combinations of cluster types which appeared less than six times were not included. Accepting frequencies of five or less, the combinations become more interesting.

Free Listing Order

Several studies demonstrate an order effect in the free listing of items from memory. It would be difficult to dredge up long lists of anything without having some mechanism to systematically recall. There is clear evidence of this effect from the free listing task in this experiment.

Caution must be taken when creating divisions in a list of names. Unlike the clustering algorithm where there are strict criteria for the formation of groups of names, the goal here is to look for logical, yet subjective, breaks in the types of people that are being listed. This is to see whether the groupings found in the clustering analysis will be found here, in a substantially easier way. Although the time of entry was stored for each name, this has proven difficult to use since informants may have paused for a variety of reasons, not the least of which is the length and difficulty of typing a given name.

In comparing this method of subgroup definition with the clustering method, remember that clusters are based upon a higher likelihood of inter-knowing among cluster members. This implies social interaction, at least cognitively, to the informant. The characteristics of the cluster members are then examined to decide why these alters would have a higher degree of inter-knowing than a randomly constructed group. Forming groups by examining the free listing order is only an analysis of characteristics. It is the opposite analysis from clustering.

The assumption is that those who are listed closely together with very similar characteristics are very likely to know one another. Although this is true in some cases, it is clearly not in others. This stems from the fact that characteristics may, but do not necessarily, reflect interaction. As an example, consider immediate family, where the mechanism of recall, members of the family, may result in a group of highly inter-knowing members.[24] However, compare this to recalling ex girlfriends. The mechanism still serves to dredge up a list of names but this group is very unlikely to exhibit high inter-knowing.

Another factor to consider which plays a less significant role is name effect. There are several cases across informants where a first or last name obviously triggered the recall of the next entry of someone with the same or similar first or last name who was unrelated by any other characteristic. Even though it was a noticeable effect, it occurred few enough times that it had virtually no effect on this analysis.

One final note of caution on analyzing the order effect is to keep in mind that this analysis came after the clustering analysis. Thus most of the groups have been determined based on the content of the cluster analysis. In the case of general and close family, this is a difficult decision. In looking at a sequence of family names, where are the group boundaries? The clustering algorithm does this on the basis of inter-knowing, here it is not as clear.

An interesting perspective derives from an examination of the first choice of each informant. In a few cases there was a clear experiment effect. One informant used me, another used his roommate who was sitting next to him taking the experiment, and another happened to know her second grade school teacher, the seemingly improbable example used in the instructions. Although these may have affected the content of their respective lists for a set of names, it appears that 60 alters was enough to force informants to cut across many groups.

Over 50% of first alters were accounted for by five types of knowing: mother (17.8%), significant other (13.3%), best friend (8.9%), father (6.7%) and sister (4.4%). Relatives in general accounted for

---

[24] Even this may not be the case. Some informants demonstrated very fractured relationships with family, including immediate family, such that there was no regular communication between members.

46.7% of primary alters. The remainder were quite varied. The most convoluted of these was the "girlfriend of a guy who was best friends with first boyfriend." It is interesting to speculate as to why this alter first occurred to this informant in the free listing task. This may be due to my instruction for alters who are simply friends. Informants were asked to at least describe how they met. This alter may actually have been a best friend, but this is how they happened to meet.

An analysis of the first and second groupings listed by informants, that is lists of characteristics that seem to hang together, do not exhibit much variability. A full 51.1% of first groups are close family, along with 17.8% of second groups. As with the primary alters, the remainder of the primary and secondary groups are rather well distributed across cluster types. This suggests a strong tendency to use family as a first choice in the free listing task, but if family is not used, there is little or no pattern.

Comparing the cluster types generated by free listing order to types generated by the ADCLUS model proved to be difficult. The inability to identify a clear cut boundary between groups for the free listing set created questionable judgment calls. In general, most cluster types found with the clustering procedure were also found from order effects. However, the membership was violently compromised by some informant's tendency to switch between types. That is, some informants alternated between family and work groups many times during their listing of 60 alters. Naturally the clustering procedure would group family or work into high inter knowing networks despite their position in the list, while they would not be grouped together for the order set given their various positions. Some cluster types, such as family including friends and general family, are very difficult to identify by order.

It would seem that a simple examination of the free list order does not produce the same detail or quality of groups as the clustering procedure

Recalculation of Clusters

As was mentioned earlier the cut off for the free list was set at 60 alters largely for logistic reasons More alters would have increased the task quadratically But what about less alters? Could the same results be achieved using 50 40 or 30?

To address this question I repeated the clustering procedure on five informants using the first 50, 40, 30 and 20 alters from their first name sorted lists This constitutes a random grab of alters from each informant's total list The results are presented in Table 27 In addition to the number of valid clusters and the number of cluster types two indices were calculated to help identify the optimum number of alters These are ratios of the valid clusters and cluster types found to the numbers of alters used

Only one case in five was able to produce clusters with 20 alters This was the one black informant of the five As with other black informants her proximity matrix was very dense at all levels of knowing making it more likely that less alters would continue to result in valid clusters All five informants were able to generate clusters at the other levels

Neither the number of clusters or the number of cluster types consistently drop with less alters Indeed in one case the number of clusters rises This is probably due to the effect of large dominant clusters which overshadow the formation of smaller clusters when all 60 alters are present

Of greater interest are the two indices which weight the productivity of the matrix i e the number of clusters and cluster types by the number of alters it took to yield them In four out of five runs both indices peak at the 30 alter level The fifth case

Table 27 Results of cluster recalculation for five informants using varying levels of alters

	First n # Alters				
	60	50	40	30	20
BETTY					
Valid Clusters	12	10	10	9	10
Cluster Types	6	5	4	4	3
Valid Clusters/# Alters	20	20	25	30	50
Cluster Types/# Alters	10	10	10	13	15
STEVE					
Valid Clusters	12	11	11	11	0
Cluster Types	7	7	7	6	0
Valid Clusters/# Alters	20	22	28	37	0
Cluster Types/# Alters	12	14	18	20	0
JENNIFER					
Valid Clusters	14	12	10	12	0
Cluster Types	8	9	7	6	0
Valid Clusters/# Alters	23	24	25	40	0
Cluster Types/# Alters	13	18	18	20	0
TONY					
Valid Clusters	11	14	13	12	0
Cluster Types	7	6	6	4	0
Valid Clusters/# Alters	18	28	33	40	0
Cluster Types/# Alters	12	12	15	13	0
MARY					
Valid Clusters	6	8	9	13	0
Cluster Types	5	6	7	6	0
Valid Clusters/# Alters	10	16	23	43	0
Cluster Types/# Alters	08	12	18	20	0

peaks at 30 for the index using cluster number, and at 40 for cluster type. Given the productivity of the matrix relative to the number of alters, and thus the task presented to informants, 30 alters seems to be the optimum level. Reducing the experiment to 30 alters would cut the length of the experiment by more than half since the final part would consist of 435 pairs versus the 1 770 pairs using 60 alters. Of course this assumes that the 30 alters are a random grab of an original list of 60 names.

The conclusion that the results can be replicated using 30 alters is based on certain heroic assumptions. First, there were only five of 45 informants who generated clusters. This may not be enough to make a definitive statement. Second, reducing the list by eliminating the hetten 10, 20, 30 and 40 alters from the first name sorted list is not the same as free listing 50, 40, 30 and 20 alters. The previous section demonstrates a significant order effect. So the subset of 30 would have to be selected from a larger free listed set to begin with. Otherwise there is a great danger of biasing towards alters who are relatives, the modal category for the first order cluster. Finally, it is questionable whether the criteria used (i.e. number of clusters and cluster types) is more important than other variables. Although cluster types are represented using 30 alters, their membership is not the same as they are at 60. Which is more 'correct' or "accurate"? This last point is the topic of the next section.

The Meaning of Clusters

Recalculating the clusters for five informants at varying levels of alters raised some serious questions with my general method. Some extreme differences in cluster membership occurred between those generated from a list of 60 alters versus one of 30. These were due to the exclusion of alters from clusters generated at level 60. Even though the representation of types was similar, the recalculation pointed to a flaw in the subjective interpretation of cluster membership.

Specifically, the problem is how to label a cluster. It has already been made clear that the subjectivity factor in assigning a cluster type to a cluster was significant. Yet the agreement of 10 informants with my interpretation of their clusters is at least partially vindicating. So informants presented with a list of people they knew agreed with my interpretation of what type of cluster it was. As it turns out, this interpretation is very fragile and highly dependent on frequencies of social relations in a cluster.

For example, a cluster was labeled a network via if it consisted of a family where the primary knowing was through one person. On a reduced list the children in the family might drop out, in which case the cluster becomes a couple. Alternatively, a cluster might be labelled family, including friends using 60 alters, but with less alters the friends might drop out transforming the remaining members into a close or general family cluster. Some examples like these occurred with the five informants, but looking at the list, the potential for many more became clear.

How can a current or former work group be distinguished from a social group, or for that matter, a hobby group from a social group? Simply because these categories were inductively defined from the results of a list of 60 alters, does that mean that they are the categories which should prevail? There are no clear criteria to decide what boundaries should be used to define relevant cluster types. Presumably such criteria would be based upon some goal- perhaps the prediction of a specific behavior.

Although the labeling appears clear to an informant, the lists themselves are an artifact of the choice to use 60 alters. Most research is impaired by the necessity to make decisions and define boundaries. However, this does suggest caution in how the results are interpreted.

Telephone Survey

While collecting these data from the recruited informants, the non-representative characteristics of my sample became evident. In an attempt to test these results in a more statistically representative fashion, I devised the following experiment for a telephone survey which is implemented monthly by the Bureau of Economic and Business Research at the University of Florida.

My objective was to randomly select a handful of alters from each respondent's global network of all the people they knew. In this case, knowing was constrained to mutual recognition (by sight or name) and contact within the past year. For each alter, interviewers asked the respondent for a textual description of how they knew them, had them rank on a scale of 1 to 3 how well they knew them and finally ranked on a 0 to 1 scale whether each unique pair of alters knew one another. Given the time constraints of the survey and the pairing routine, the list of alters was set at eight.

To select eight people from a global network it was assumed that first names were randomly distributed. In other words, there is no reason to believe that Johns are more likely to be work associates than Joes. However, there could be a gender bias. It is possible that Johns are more likely work associates than Janes although less so today than in previous years. For this reason a male list and a female list were drawn.

Another factor, which hopefully was addressed by the source of the list was name trends. Some names are fashionable at certain times. For instance, Stephanie is a popular girls name now while Michael is popular for boys. Thus there could be an age bias if these are not distributed randomly. The names were drawn from a mailing list maintained by a research center at the University of Florida which presumably includes all ages above about 20. Granted those under 20 could possibly be biased against, although common all time names are still represented and the survey is answered by respondents over age 18.[25]

Interviewers were first presented with a random version of the male list. They read off names until the respondent acknowledged that they knew someone with that first name. When respondents recognized a name they described how they knew them and ranked on a scale of 1 to 3 how well they knew the alter. They were then presented with a female name list and provided the same information. This alternating presentation continued until four male and four female alters were collected. Respondents were then asked how they applied the values of 1 to 3 in their ranking of alters. As with the extended instrument, the unique pairs, 28 in this experiment, were presented for evaluation of mutual knowing.

---

[25] There are a few other name biases as well mostly concerning ethnic names. For instance Hispanics or Asians are very likely to have ethnic names for family members. The list that was used has a distinct European bias as is typical of American first names. There are examples of Hispanic names, but not many. This was not felt to be a serious problem since the number of Asians in Florida is not significant, and since Spanish interviews (that is, interviews for monolingual Spanish speakers) are restricted to the permanent questions of the survey, and do not include one time research questions which is the category these fall into.

Another criticism of this method has to do with the selection of only eight alters. For those with clusters of widely varying size or many cluster types, theoretically this would bias selection away from small but potentially important clusters. For example, most academics have large pools of potential alters from professional associations. While family clusters may be small relative to these, in many ways family clusters carry more weight. That is they probably affect more behaviors than do professional associations.

A large professional association, given its large size, is more likely to contain any one name than a smaller cluster. Similarly, respondents with many clusters are more likely to have a name that fits a non-family cluster, since there are more of them, than respondents with fewer clusters. However, if the percentage of family clusters found is any indication (76%), this was not a problem for the survey.

Coding the textual data for the telephone survey proved to be less accurate than the extended experiment described above. This is largely attributed to three factors. First is the collection of only a textual description for each alter. Without the other data to corroborate coding, the assignment of values became difficult. Second was the fact that undergraduate interviewers were intermediary between the response of the respondent and the computer. This in conjunction with the absence of monetary incentive for the respondent, resulted in less detailed, and sometimes suspect, responses in some cases.

To accommodate the loss in accuracy, broad categories for coding assignment were used which were not as necessary with the extended instrument. For instance, one respondent labelled an alter a relation from a former job. Another respondent merely said 'work' or that is all the interviewer typed. Since it was difficult to judge whether 'work' meant current or former, a coworker, boss or employee, a general category called 'work' was created. To use these responses, they became the least common denominator and all detail is recoded as 'work'. I cannot argue the case that 'work' be contrasted with detailed responses. The categories found reflect those found by Fischer et al (1977), and are similar to the collapsed categories from the extended instrument.

The results of the survey are presented in Tables 28 and 29 and

Table 28. Percent of informants who mentioned category, by three levels and over all levels.

Category	Level 1	Level 2	Level 3	All Levels
Frequency of contact	14	29	16	38
Friendship	2	11	22	29
Closeness	10	8	9	26
Situation specific	13	16	7	28
Family or relative	4	5	16	24
Emotional aspect *	19	9	5	22
Acquaintance	19	3	0	21
Amount of information known	0	2	23	20

Note: Includes descriptions of 'casual' relations.

should be compared to Tables 5 and 20 respectively. Table 28 shows the knowing themes expressed across the 233 respondents who completed the survey. Notice that the categories do not match directly. The difference is the inclusion of situational specific and conversation content (strongly represented by the concept of casual relations) in the survey coding. Otherwise the categories are the same. With the exception of the closeness category, there is a very similar pattern to the usage of these themes by knowing level. For instance, acquaintance

Table 29. Percent of informants exhibiting knowing categories for male, female and all names combined

Category	Male Names	Female Names	All Names
Relative	54.1	57.9	76.0
Work (current and former)	57.1	57.1	71.7
Network via another person	42.1	42.9	63.1
Neighbor	21.9	22.3	34.3
School (includes teachers)	12.9	17.6	23.6
Just a friend	9.0	10.3	15.5
Provides a service for ego	9.4	6.4	13.3
Religious affiliation	6.3	10.7	12.9
Social activities	8.6	5.6	11.2

seems to be a concept that is generally associated with lower knowing levels and not higher ones, where friendship and family are the reverse. Coding closeness from the survey data was slightly different than the extended instrument as it also included other emotional relations.

There are some differences in the results between the experiments that are reflected in the coding. The survey method generated several situation specific themes. For example, a respondent might have assigned a 2 to someone who they would have a drink with after work, or a 1 to someone from work.

In general, the percentages of informants expressing a theme are similar relative to their order. Friendship and frequency of contact flip flopped in the survey analysis although they were very close in percentage in both analyses (in placement, but not in magnitude). The only dramatic difference is amount of information known which was an important concept in the extended instrument (55.3% of alters mentioned it). The survey results show only 19.7% of respondents mentioning it.

In comparing Table 28 to Table 20, it is important to note that Table 20 is based on cluster types, while Table 28 is based on relation types.[28] Thus some categories in Table 20, such as 'Couples', do not make sense in defining a dyadic relation. Otherwise, the results of the two tables are comparable. Looking at the percentages for male versus female names, there appears to be little difference. This argues in favor of treating them the same.

Referring to the nonage/binary data, the order of categories, from highest percentage representation to lowest, is much like Table 20 although the magnitudes are different. This serves to validate the results from the extended instrument in terms of their representativeness, assuming the reader is willing to accept the long list of caveats detailed above.

The one notable difference is the importance of school relations, which is much lower (by half) in the telephone survey experiment. As was mentioned before, my sample for the extended instrument drew heavily from those with a university background, understandable for a university town. Thus the importance of school relations was unduly high.

A correlation analysis of knowing themes and relations yielded mixed results. The analysis of themes corroborated two correlations from the extended instrument, that of friendship and family (.35 at .01) and acquaintance and friendship (.22 at .01). The other correlations from the extended experiment involved categories that were not well represented in the survey experiment.

Surprisingly, the correlation analysis of the knowing relations yielded several significant correlations, all negative. My conclusion was that these were nonsense. The consistent negative relation was due to the small N for each respondent. Eight alters tends to be too small to cover many clusters, thus creating a bias against cooccurring relations

---

[28] Early on in the telephone survey, an added step was in place to elicit cluster membership. Respondents were asked to think of two people they and the alter knew in common and to identify the group to which they all belonged. This is based on the theory that it is unlikely for four people to know each other unless they are part of a group. This section was removed as it proved too time consuming for the telephone interview, and much of the information was redundant with the textual description of the knowing relation. However, those results that were collected were promising.

Network Density

Among the many measures applied to network matrices density is perhaps the most common and simple Assuming a symmetric 60x60 matrix there are 1 770 possible ties excluding the diagonal which signifies an alter knowing him or herself Network density is the percent of measured ties in this case the number of ties divided by 1,770 Since each of the 45 informants with usable matrix data has a different matrix there are 45 density measures

As was mentioned before informants could rank knowing between alters on a scale of 0 through 5 The final analysis of the clustering procedure was applied to matrices which were recoded such that knowing was defined as a ranking of 4 or 5 Table 29 shows mean density for various levels of recoding Not surprisingly the density drops as the criterion for knowing becomes more stringent Mean density where a tie is counted for any non-zero ranking is significantly (001) higher than the mean density where ties are counted for rankings of 4 or 5

Table 30 Average densities for 45 informants for various recoding schemas

Density Criteria	Density	SD
Value 1 to 5 = Tie	27	09
Value 2 to 5 = Tie	18	09
Value 3 to 5 = Tie	.14	.08
Value 4 to 5 = Tie	.08	.05
Difference between 1 to 5 and 4 to 5 as a Tie	15	08

Running t tests using the least and the most restrictive measures of density as dependent variables resulted in some unexpected results which are displayed in Table 30 Although there were only four black informants they had significantly higher mean densities at both levels Blacks tended to use a high percentage of family and neighbors which accounts for many of the ties Married informants had significantly higher densities than unmarried informants with the more restrictive tie definition, but not with the lower This again may be due to the presence of more family from the spouses side, which is a large high density group Non-family are most likely lower density groups in the case of both married and unmarried informants showing up as low density ties equally between married and unmarried informants

Table 31 Results of t-tests on two density measures between selected groups

Group	N	Mean Density (SD)	T value	Prob > T
RACE				
1 to 5 Tie				
Black	4	41 (12)	3 01	.05
White	41	22 (07)		
4 to 5 Tie				
Black	4	18 (08)	2 58	08
White	41	07 (04)		
MARITAL STATUS				
1 to 5 Tie				
Married	16	24 (08)	-0 15	88
Not Married	26	23 (09)		
4 to 5 Tie				
Married	16	10 (06)	-2 10	05
Not Married	26	07 (04)		

At least one of Granovetter's weak tie propositions could be tested with these data According to Granovetter's theory, alters who are weakly tied to ego will demonstrate a matrix of lower density (lower percentage of ties) than those who are strongly tied The question arises as to what constitutes a weak tie. Some might argue that level 3 constitutes a strong tie while others would not

Table 31 presents the results of a weak/strong tie analysis of these data Here weak ties have been defined as alters who informants rated with a knowing of level 2, 2 or 3 and 2, 3 or 4 Since no informant rated any alter with a knowing level of 1, this was excluded In all three cases subsets of weak tie alters demonstrate significantly lower density than subsets of strong tie alters This lends credence to Granovetter's proposition

---

Table 32 Results of t tests between weak and strong alters given varying definitions of weak/strong tie

Weak ties <= #	Average weak density	SD	Average strong density	SD	Prob>T
2	11	16	17	09	03
3	07	06	17	01	01
4	03	03	24	17	01

Upon closer examination, however the average densities exhibit a curious pattern As expected, exclusion of lower strength ties from the strong tie definition results in higher density, more ties on average However, this does not happen until strength 4 alters are excluded from strong tie definition More curious is the steady drop in weak tie density as stronger tie alters are introduced. If Granovetter's proposition were true outright intuitively one might expect the inclusion of stronger ties to the weak tie subgroup to raise the density

To the contrary, the explanation for this further supports Granovetter's theory Assuming that adding informants of a higher knowing level constitutes adding an interacting subnetwork whose members are more than likely not tied to their lower strength neighbors, the addition of new members can easily add more potential ties than actual ties. This can, but does not necessarily, result in a lower overall density even though the added submatrix is highly interactive.

Network Degree

Network density indicates how interconnected the network of 60 alters is as a whole. Clearly though, some alters are more connected than others. Measuring how many connections an informant's alters have on average paints a different picture from network density since some alters contribute more to the density than others. This measure is known as the "degree" of a network.

Defining knowing as a tie of strength 4 or 5, alters had on average 5 ties (SD 5.6) out of 59 possible. An examination of Figure 2 as well as the standard deviation, reveals the wide variability between informants on this statistic. Some alters knew only the informant while a few alters knew as many as 44 other alters in the informant's matrix. Forty four ties is a sobering number, suggesting that these alters approach the connectedness of the informant, without the advantage of selecting their own list.

As with network density, black informants exhibited higher degree than whites. The average of 11.0 (SD 8.8) for the black informants was significantly (.001) higher than the 4.5 (SD 4.8) average for whites. Blacks not only appear to have more integrated networks than whites, but highly integrated members as well.

Who are the highly integrated alters? To address this question I selected each informant's highest tied alter. In the case of ties (i.e. more than one alter with the maximum number of ties), both were counted. Since there were three cases of ties, a total of 48 alters were subset

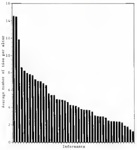

Figure 2. Average number of ties per alter for each informant

Sixty seven percent of the 48 highly integrated alters were family. These included parents (23%), significant others (17%), siblings (10%), children (10%) and grandparents (6%). Another 17% were work related alters and the remaining 17% were miscellaneous. Again this may be a circular observation. Since the modal knowing type for all alters was some type of family, and since family clusters are highly integrated, it is to be expected that the most highly integrated alters would be family.

CHAPTER 6
CASE STUDIES

Introduction

Analyses such as have been presented in the previous chapter are often illuminating, but frequently suffer from the alienation of the individual from the statistics. For this reason I have selected five informants for description of their particular cases relative to these statistics. Hopefully this will make clear any data manipulations which were difficult to understand on an aggregate level. Additionally, these should show the way in which the measurements can be applied to individual cases.

The shortcomings of both aggregate and individual analyses come to bear in this examination. Although something is certainly lost in the aggregate analysis, selection of particular cases is even more difficult. Among the 45 informants whose clusters could be used, there is wide variation in backgrounds and results. The variety of experiences each has had within this sphere alone accounts for vast differences in the structure and composition of their networks. Therefore, any set of five is not likely to represent the whole. In order to make sense out of both chapters 5 and 6, they must be considered in tandem.

My selection of these five cases over others was in large part based on my personal knowledge of these informants. Simply put, there was not nearly enough information provided by the experiment itself to understand the intricacies of most informants' networks. There were informants whose networks were interesting, and provided radically different cases from these five but they could not be interpreted adequately.[27]

Steve

Steve's case history highlights the effects that location plays in the formation and maintenance of subnetworks. Originally from Australia, Steve pursued an MBA in Toronto, Canada and is currently working on a PhD in Gainesville, Florida. He is best characterized as an energetic person who is actively involved in a variety of groups. For instance, in Gainesville Steve maintains contact with both his graduate cohort and certain faculty members. His membership in the Sailing Club, Water Ski Club and Hash House Harriers (a running club) provide him with a variety of avenues for social life and entertainment. He devotes some of his time to helping out in a shelter for the homeless. His exuberant and gregarious nature seems to drive him to seek new groups with a seemingly endless energy.

A general examination of his list of alters shows similarities to the group in terms of education. Yet he listed more female alters than males did on average, 52% to 43% respectively. Steve perceives more loyalty to his alters than most informants in response to the question whether a change in location would terminate the relation after a year. Only 3 of 60 (5%) would not be known, he says, compared to 29.4% across the entire sample. Although it is difficult to say whether this would be true, Steve has demonstrated this loyalty in the content of his alter list—it contains many people from other countries who he has not seen in years. It is also interesting that the mean age of his alters is significantly (.001) younger than the average (31.4 compared to 36.6) and that the number of years since his last contact is significantly (.04) lower than average (.8 compared to 1.2).

---

[27] One case in particular, a devout Hare Krishna follower who migrates seasonally between Gainesville and New Jersey doing work as a truck driver, had enormous potential for an interesting case.

Oddly though, a test on the maintenance variable shows him to have significantly (.02) more type 4 relations, that is, situations where effort is not expended by either person to maintain the relation. This would appear to run counter to the locational aspects of his network.

It is not unusual that Steve generates more sensible clusters than most informants, 12 clusters compared to an average of 9.58. However 12 clusters is well within one standard deviation (2.83) from the group average. He also used slightly more cluster types than the average (7 compared to a mean of 6.33) although again within one standard deviation (1.82).

Upon examining his clusters, they appear very diverse geographically. This, in part, explains why the overall density in his network of 60 alters is lower than average. Using the 1 to 5 recoding his density is .16 compared to .27 (SD .09) for the group, which is more than one standard deviation from the mean. The 4 to 5 recoding yields a density of .05 compared to an average of .08 (SD .05), within one standard deviation. This suggests that in a more general knowing scheme he is more diverse than most informants, involving himself in groups across locations and types. But the more restrictive coding is lower, yet not demonstrably so, than most informants. His close of people who know one another very well are similarly linked as those of other informants.

Steve's cluster types reflect his involvement in a PhD program. Of the 7 cluster types used, university accounted for five clusters. On the other hand he had no close family clusters at all. This does not mean he has no close family; indeed he does. However, there is not a strong core of interacting family as with some other informants. Steve has verified this suggesting that his mother is a focal point for he his brother and sister, but there are few occasions where several family members get together. This makes it difficult to maintain high interknowing family relations.

The strong presence of hobby and social groups comes second only to the high number of university related groups. Note the negative correlation in Table 22 between close family clusters and hobby clusters. Steve reflects this relation perfectly. In the absence of interaction with close family, he seems to have many hobbies and other activities which provide his close relations.

This reflection of dispersed and unrelated groups is perhaps most noticed in the area of overlap. Steve is below the average with only 2 of 60 informants crossing cluster types, compared to 6.36 (SD 4.65) for the group as a whole, barely within one standard deviation. Given the low number of overlaps and the low overall density, but the higher number of sensible clusters and the high number of cluster types, Steve's global network seems to be made up of many concentrated groups with little crossover. This fits his mercurial social strategy, moving from place to place, group to group, retaining cores of alters.

Figure 3 is a plot of the MDS coordinates for a two dimensional solution. These were run with the ego to provide some sense of the informant's relation to the group as a whole. It is obvious that Steve is well centered in this group. Finding some underlying dimensionality is difficult even though his clusters are easily recognized. The lower right consists of Gainesville alters, concentrated around hashers (fellow runners). The upper right presents the Toronto MBA alters. The left side of the plot appears to be made up of Australian clusters. If there is any dimensionality in this figure it would be based on location.

Betty

In stark contrast to Steve who is white, male, academically oriented and highly mobile, Betty is a black female university secretary with a high school education and strong ties to her home area. Unlike Steve who has travelled extensively, Betty has rarely travelled outside

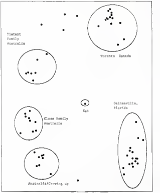

Figure 3. MDS plot for Steve

of Florida and Georgia She has a strong sense of belonging to the area and, as will become clear has close family ties

Having invested several years in the university as a secretary, she is typical of many clerical personnel in the Florida state employee system These are people who are not mobile given their qualifications (usually a high school education) and their strong family ties To Betty, work is something you do to make a living and not a high valued life experience

Her private life is separate from her work life A strong relationship with her husband dominates her lifestyle such that any participation in social groups centering around work is merely out of obligation She visits her family in Georgia, particularly on extended vacations

Betty names more females than males (62% to 38% respectively) in her network many more than the 56% female informants named on average This may reflect an overall dominance of women in her networks The educational level of her alters is close to the average in all categories In terms of maintenance of the relationship, Betty has no examples of situational relations a significant (001) difference from the group average indeed, 83% of her relations are seen as equal effort between herself and the alter, with the remaining 17% maintained by her

Her mean knowing level and mean closeness to her alters are both significantly (001) higher than the average over all 2 820 alters at 4 1 compared to 3 5 for knowing and 4 5 compared to 3 1 for closeness Oddly she would lose contact with a much higher than average percentage of her alters than the group would if she moved away I would have expected high knowing levels to be associated with those whose social relations would prevail over distance Betty presents the image of someone who knows someone well when in spatial proximity, but has little intention of maintaining contact after moving

The number of years since she has had contact is significantly (.001) lower than the group Network density at both levels is much higher than normal, particularly at the restrictive level where it is three times the group average

Betty had a total of 12 clusters and seven cluster types, both within one standard deviation of the group averages Her current work accounted for half of the clusters while close family accounted for two She failed to include her husband on the list despite his apparent important role in her lifestyle

By all appearances Betty's network is dominated by work although this could have been a protective mechanism on her part to avoid divulging personal information She is very close to her alters This may be in part due to her personal interpretation of the scale Someone who is relatively assertive about their life would be expected to reserve high knowings and closeness for very special cases That is not the case with Betty Many work alters received high scores for knowing and closeness In fact her definition of the knowing levels lists category 4 as "Work closely with or socialize " Five is reserved for good friends or family Since this is an ordinal scale this raises the question as to how much distance there is between her 4 and 5 ranking compared to her say 3 and 4 My suspicion is that it is larger

The high density is explained by the dominance of two primary cluster types current work and family The high restrictive density is due to the unusually high scores associated with work alters which make up such a large portion of her alter list

Unlike Steve Betty has a high degree of overlap between clusters and across cluster type However these are almost completely accounted for by the overlapping of work and social group clusters -an anomaly considering the dominance of family in overlapping cluster types When Betty is social it seems that it is with people from work Again, this conclusion is drawn under the assumption that the 60 people she mentioned are the people in her global network with whom she interacts the most This is very likely the case since Betty's family lives in Georgia

Figure 4 again demonstrates the importance of work relations in Betty's list of 60 alters The right side of the plot is completely dominated by her current work Far on the left is her family which shares no links to her work, and her in laws at the bottom who are equally separated from her biological family

Tony and Mary

Tony and Mary are a couple who present a very interesting case Having met in high school and dated for many years they finally were married when they graduated from college Neither has ever seriously dated anyone else

Tony received a degree in business administration and computer science and Mary a degree as a dental hygienist They moved to the Orlando area after graduating from the university and worked as a two income family for several years Tony built up a career in the computer division of a large company while Mary became very successful as a hygienist for a dental practice Both Tony and Mary had family in a town very close to Orlando and had several groups frequently couples with whom they socialized They have no children

Four years ago Mary decided she wanted more in her career She started taking preliminary courses towards entering dental school and was eventually accepted to the dental program at the University of Florida from which they both graduated They moved back to Gainesville where Tony accepted a position which paid substantially less than he made in his career track in Orlando Faced with a single income which was less than either earned individually they were forced to get a roommate to alleviate the financial burden

Dental school is quite rigorous, particularly in the earlier years The 12 to 16 hour days leave little time or opportunity for

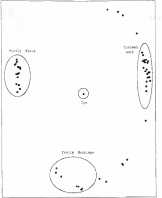

Figure 4   MDS plot for Betty

establishing relationships outside the dental school routine. To further aggravate this, Mary tends to be very ambitious, thus spending even more time at the dental school than most students. This leaves less time for socializing with Tony. Tony, in many ways, is the opposite case.

Like Steve, Tony is a gregarious person who enjoys socializing with others. Although he may spend some time at home, he enjoys getting out. Whenever Mary is available, he spends all of his time with her and because of her obligations to socialize with the dental organizations, he attends many of their functions. Yet these dental functions are not an everyday occurrence. To satisfy his social needs Tony developed many network groups independent of Mary.

Mary used exactly half male and half female alters, which is a much higher percentage of males than female informants used on average. In terms of education, Mary's alters tend to be mostly university and above. This, of course, is accounted for by the cluster of dental students, close to half of her entire list of alters. Intuitively, this would seem to be a structure where most of the relationships are seen as ascribed, that is, the relationship is situational. Instead, Mary perceives most of her relations (80%) as mutual. Her network tends to be more dense than the group on average.

Her less restrictive density (any knowing recoded to 1) is .33, nearly 10% higher than the group average of .24. More interesting is her restrictive density which is well over double the group average (.2 compared to .09). This is more than two standard deviations from the group. The high level of less restrictive density is explained by the dental class. This is a pool of alters who all know each other.

The high value for the restrictive density is curious, it is not the case that most of her high knowing alters are dental students, although some are. Indeed, it seems that her high knowing alters cross groups. Perhaps compared to other informants, her high knowing alters constitute several subgroups rather than high knowing individuals that exist in many groups.

It is not surprising that the incidence of college educated alters is significantly (.001) higher in Mary's list than the rest of the group. The average number of situational relationships, knowing and closeness are not significantly different from the group. Age and the number of years she has known her alters are significantly (.001) lower than the average. In no case has it been more than a year since she has contacted an alter.

With six sensible clusters, Mary is over one standard deviation below the group average, yet with 5 cluster types she is close to the group average. This is again due to the utter dominance of the dental school cluster which explained 69% of the total variance in the cluster analysis. There is little room for the expression of other clusters as the dental cluster accounts for over half of Mary's alters and there is a high degree of interknowing within that cluster. But those that are expressed are small and of varying types.

In sum, Mary's list of 60 reflects the overwhelming influence of dental school on her life. This is a highly interacting group that overshadows any other potential area of social relations.[28] Notwithstanding, there is evidence of other groups, mostly those she knows through Tony or which were hers prior to her entrance into the dental program. Her single family cluster is a mixture of close and distant family, with no tight knit core. Dental school has left little time for a social life outside of the dental group.

Tony is close to the male distribution of alter gender, using more males than females. In fact, Tony is very close to the group average on education level the location where he met his alters the number of alters he would not contact if he moved knowing levels, closeness levels and both types of density Other than the years since last contact which like Mary was 0 in every case[28], the only variable Tony seems to differ on is the maintenance of relationship He has no cases where someone else maintains the relation but significantly (.001) more incidence of situational relationships more than double the average

---

[28] Pool and Kochen (1978) suggest that there are finite limits to the number of network ties a person can maintain. This is sensible as maintenance of a network requires time and effort, some ties requiring more effort than others. Thus, Mary's involvement in high maintenance, highly dense relations, i.e., dental school, drew her away from other clusters (see Wellman and Berkowitz, 1988, pg. 42).

I initially explained this as a misinterpretation of the use of the category since nearly 10 alters are considered close family and are viewed as situational knowings However there is selectivity in other categories, such as work In applying this That is, some work alters are considered mutual relations others are situational In turn some family are mutual, others are situational Tony explains the close family situational phenomena by family reunions and holiday gatherings They are family which he grew up with but now only sees when visitation with very close family creates the situation The situational relations are always cousins aunts and uncles Most of the other situational alters are relations through others such as someone s spouse

The analysis of Tony's specific clusters reveals 11 sensible clusters, which is close to the average but eight cluster types which is nearly double the group average Tony's clusters include strong family representation, with close, general maternal and paternal clusters This reflects the values of his rural upbringing in Appalachian Kentucky Tony's parents are very young relative to Mary s who are older and immigrants from Eastern Europe Mary s parents retired to Florida very early in her childhood Thus she may not have family available to form clusters while Tony does

---

[28] Neither Tony or Mary had any trouble listing their 60 alters The fact that they have had contact with all of them within the past year is evidence that they tend to either know many people, making it unnecessary to reach back in the past for their alters or they are conscientious about keeping their social relations up-to date

In sum, Tony appears to reflect the group average in almost all categories except the incidence of situational relations and number of cluster types. He has strong family and similarly strong representation from work and former work. Indeed, it seems that wherever Tony worked he has a cluster of people he maintains. These very likely were and are social groups he has moved within for many years.

Mary and Tony's MDS charts are interesting (see Figures 5 and 6). On both charts they are each other's closest alter, directly proximate with respect to everyone else. In Mary's chart the dental cluster with 40% of her alters, the Orlando grouping and the grouping of business associates are easily distinguished. Even though she is highly embedded in the dental cluster, her relations to the other clusters involve Tony. In fact, Tony is somewhat integrated into the dental cluster by virtue of his attendance at the many dental school socials where spouse participation is encouraged. Similarly, Tony's clusters of family, work and former work are equidistant from he and Mary in the center. She is a member of the clusters as well, largely through him.

In sum, Tony and Mary appear to be cluster brokers for each other. She provides linkage to one large and dominant cluster, the dental school, and he for several clusters, mostly family and work related groups for her. Nevertheless, they include each other in their disparate social spheres.

Jennifer

Jennifer is a white female in her late twenties who was born and raised in Tennessee. She went to school in Florida and later moved to Gainesville where she got her first job. A graduate level economist, Jennifer early on had aspirations to a higher paying and more challenging position. This has led to her current position which is notable for a person so young.

For a long time Jennifer disliked Gainesville for what she perceived to be a low quality social life. Strong ties to family and

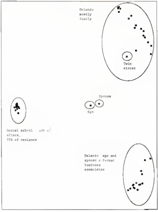

Figure 5 MDS plot for Mary

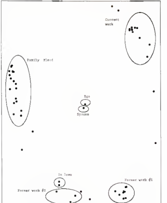

Figure 6. MDS plot for Tony

feelings of starting a family of her own, despite her career goals pushed Jennifer to find a partner who would be willing to share her need for family. Being a college town, many of the men Jennifer met were not of this type. She thus became relatively disaffected from the social groups to which she attempted to become integrated.

Through a friend, Jennifer was introduced to her husband Brad, who was also an informant in this study. Within six weeks, Jennifer and Brad decided to become engaged. This introduced a new source of network alters for Jennifer since Brad was previously married and had a daughter. This serves to demonstrate the fluidity and dramatic changes personal networks undergo.

Jennifer, like Tony, follows the average for her own gender in terms of the sexual distribution of her alters. Her alters tend to have significantly (.001) higher education than the average, but like Tony significantly (.001) more are situational relations than average. What is more, she would lose contact with a very high number (40%) within a year after moving from Gainesville.[28] Both measures of network density are close to the group average.

The measures of knowing and closeness are not significantly different from the average, at 3.4 and 2.9 respectively. Her alters are significantly (.009) older than average at 41.4 (SD 14.2). Again, like Tony and Mary, these are alters that she has known for significantly less time (.001) than average and the number of years since their last contact has been significantly less than average.

On average, Jennifer has many more clusters (14) than most informants. Similarly, she has more cluster types, with 10, than average. The majority of the clusters are accounted for by work or former work (6 of 14) Her current work alone accounts for four clusters This is not unusual given her position which affords visibility with a variety of groups and participation in work projects where staff are forced into higher interacting subgroups

---

[28] It is worth repeating that negative responses to this question very likely carry weight while affirmative responses may not. In other words, an informant may have all intentions of keeping in contact with an alter but due to unpredictable circumstances may not. But an alter who is currently known, but is anticipated not to be known is very likely to remain that way.

The MDS results (Figure 7) show a clear dimension of family versus work On the right the clusters of current versus former work contrast one another Unlike Tony and Mary her spouse is not as close on the plot This of course does not mean that she and her husband are not close but that her husband is not as fully integrated into her subnetworks as Tony and Mary are in theirs As further evidence of this Jennifer is placed in a similar relation to Brad on his plot This is explained by the amount of time Jennifer and Brad have known each other and the limited opportunity they have had to become integrated into networks that previously belonged to them individually

Jennifer presents a case of an informant with strong family ties who had been striving for a sense of belonging in some social context in Gainesville Her marriage to Brad was relatively recent at the time of data collection and their meeting and engagement were not much prior to that Thus there is still a sense of development of their participation in each other s networks and in mutually developed networks Her work has clearly filled a large part of her life in terms of the people she knows

Selected MDS plots

As an analytical tool MDS plots are very useful although there is strong argument as to their value in determining well defined group boundaries Nevertheless it is useful to examine an additional eight plots to get a feel for the variability in the grouping of the 19 cluster types

Figure 8 is an MDS plot of my set of 60 alters As with most of these plots ego is centered among the various cliques The clique types reflect my work, my involvement in two academic departments in the

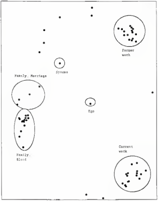

Figure 7 MDS plot for Jennifer

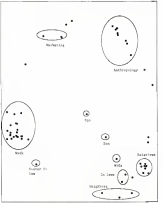

Figure 8 MDS plot for Chras

university where I work, and my family life. Perhaps most interesting is the lower right quadrant of the plot. The clique comprised of my family is tight due to the high interknowing and communication that occurs within my immediate biological family. However, the group formed by my wife, son and in laws, and to an extent our neighbors, is actually closer, although it forms a diffuse group. This is due to the fact that none of my family knows my in laws and neighbors other than a brief encounter with my mother and one brother.

My current work forms a clique apart from all other groups. Socially and practically, this unit is autonomous from other departments in the university and within the specific college where I work. Thus there are few links to anthropology or marketing.

Figure 9 presents the MDS of Arlene, a graduate student in anthropology. Her map shows three truly identifiable groups: work/graduate school, hobby and family. ADCLUS produced more detail in the family and work/graduate school clusters, detail which is lost in this plot. Most of this detail occurred in the work/graduate school area. Her work at the university museum overlaps with her graduate work since many archaeologists, her major, work in the museum and have faculty status in the anthropology department.

Brad, a commercial artist who works for the university, generated the plot in Figure 10. Of immediate note are the cliques of his ex wife's family as well as his wife. Brad is married to Jennifer from the above case. Recalling that Brad and Jennifer have not been married long, it is clear that Brad's former in laws play a role, albeit a negative one, in their lives.

David is a professional researcher at the university. Figure 11 suggests a very compartmentalized network where David maintains three main subgroups: work, family and school. In fact, the school clique may be an artifact of the length of the free listing task since many of these appear in his final 20 choices. It is questionable whether David

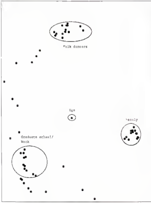

Figure 9 MDS plot for Arlene

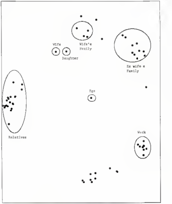

Figure 10 MDS plot for Brad

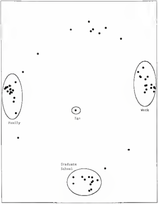

Figure 11. MDS plot for David

actually maintains these contacts as if they were an attempt to fill out a list that had already been drained of active members by choice 40 Again many informants expressed frustration with the task of listing 60 active network members

Figures 12 and 13 are MDS plots for a married couple, Glen and Kim This is a young couple who have been together for several years however like Tony and Mary they have different family backgrounds Glen comes from a large extended family many of whom reside in Gainesville This extended family includes aunts, uncles and several cousins These are not only relatives but people with whom Glen socializes frequently Kim comes from a small family Her parents divorced when she and her brother were very young Thus she has few close family ties of her own

Both plots show Glen's family as a highly populated and dense group There is no disputing the effect these people have on both Glen and Kim Indeed other than a few relations with some nursing school members Kim's network is more or less dominated by Glen's which is largely tied in with his family They know couples who they socialize with many who are ex-roommates of Kim's, but in Kim's own plot they are lumped in with Glen's family

Figure 14 is a plot of Janet, a staff assistant in a research center at the university Like many of the clerical staff, she was born and raised in a small town outside of Gainesville She like much of her family has remained in the area her entire life Thus much of her network is dominated by both her and her husband's relatives having been married for many years her husband is centered between these groups along with her This is similar to Mary and Tony who have also been married for many years Contrast this to Jennifer and Brad or Kim and Glen who have not yet developed this level of mutual network integration Janet's work and home town also account for small cliques although they pale in comparison to the family cliques

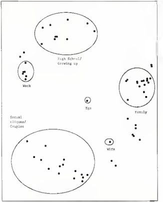

Figure 12 MDS plot for Glen

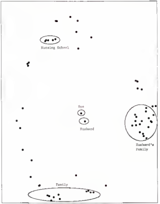

Figure 13. MDS plot for Kim

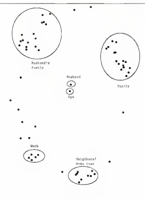

Figure 14. MDS plot for Janet

Finally, Figure 15 is an MDS plot for Pam, another black informant. Like Janet, family fulfill a major role in her set of social relations. Indeed, her use of family, friends and neighbors from a small Florida town made it very difficult to establish clear clique boundaries. The appearance is of a very dense (.47 using all knowing levels) and interactive group and few subcliques. Whether Pam truly perceives these large cliques is unclear.

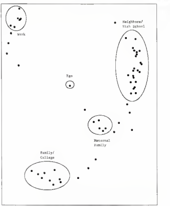

Figure 15 MDS plot for Pam

CHAPTER 7
DISCUSSION AND CONCLUSIONS

Experimental Method

As might be expected, the conduct of this experiment uncovered several flaws which could be corrected in future replications Although both the extended interviews and the telephone survey method were pretested it was not until the full analysis was underway that some problems became evident The extended interview method will be discussed followed by the telephone survey method

Recall that the method for the extended interviews was to generate a free list of 60 alters collect information about them and have the informant rank the perceived knowing between all the unique pairs

Given the utility of the resulting adjacency matrix, the last part that is the perceived knowing between pairs appeared fruitful The random checks content and meaningful clusters and checks on clusters suggest that this task yielded a rich matrix Since the length of the task increases quadratically with an increase in the alter list 60 alters is close to the maximum Several informants in fact had difficulty listing even 60 alters despite the rather loose criteria. This suggests that 60 alters produces adequate coverage of global networks[31]

However the results of the recalculation of clusters on 30 randomly selected alters from five informants suggests that smaller results can be obtained with a much smaller list reducing the time of the overall experiment Such a reduction in the tasks would make it possible to get more informants as the duration of the experiment would be reduced from three to four hours to less than two In terms of informant tedium the biographical data gathering task was more stressful than the paired comparisons

---

[31] Recall Pool and Kochen's (1978) assertion that any given person's network ties are not merely finite, but much smaller than is intuitively thought This I feel relates to limited resources for maintaining ties Network ties are not free there is an expenditure of resources to maintain them Whether 60 is the upper limit of most informant's network is a function of a reasonable definition of knowing

I found the 0 through 5 scale to be ideal informants seemed to appreciate the utility of a wider range of numbers from which to choose than if the scale had been constrained to 0 through 3 In the analysis it appeared that this forced informants to think harder about who would get a higher value It is possible that a seven point scale would have yielded even more distinction[12]

My biggest change to the method would undoubtedly be the collection of information Although a large part of the data such as the demographics remain to be analysed they took too long to collect given their utility The most useful data for identifying subgroups from cluster output was the textual description of how the informant knew the alter Secondary was the structured coding

Substituting an extended textual description for both the structured coding and structured demographics would be more useful in identifying subgroups This might help in the definition of subtle subgroups which appeared meaningless given the quality of the data This would, again, reduce the time of the experiment However I would more likely type this information myself, interviewing the informant allowing the flexibility to probe for more detailed information The analysis was constrained several times by incomplete descriptions

Although I did a check of a subsample with positive results the participation of more coders would only increase the reliability of any findings This would eliminate any bias I may have had towards deciding what should and should not be called a valid cluster

---

[12] This assertion is tenuously supported by Miller (1956) who attributed special psychological properties to the number seven

Finally I would like to get a better representation of the general population than was achieved. A shorter instrument would make this possible Assuming the reader accepts the results as comparable the telephone survey suggested a strong bias towards school clusters in the extended interviews since they were not nearly as prominent a category in the survey As my general thesis suggests, who you know is not random nor who they know My sample was closely associated with the university or community college To get a better idea of the groups that exist some form of random selection on a larger group would be better Indeed, cross cultural comparisons might show more variability

Several of the faults of the telephone survey method were discussed in Chapter 4 However how to improve on the method is another matter The advantage of a telephone survey is clearly the reliability it provides over an extensive small sample experiment where validity is maximized There are elements of a telephone survey that will always be problematic In general first names as a cue work quite well It is a tag for a person that is known for most significant social relations although clearly not all, and it is generally randomly distributed with the exception of the name biases The limitation to eight alters actually biased against group formation Unfortunately with a telephone survey the paired comparison test can only be done on a very small number of alters Thus it is questionable whether this is a good method for establishing subnetworks over the telephone

In terms of identifying types of social relations, the method was useful The limiting factor for this was the interviewers themselves Since several people were involved, with various ways of interpreting answers and varying levels of motivation the resulting descriptions had to be generalized to a great degree Greater accuracy, and possibly more significant results could be achieved by using closed categories

However, the results of the extended experiment verify the assertion that these categories are not mutually exclusive. Alters can be coworkers and attend the same church. Thus a closed list where the informant can pick up to five descriptors of the relation might be better. To implement this, the precoded categories must be known. More than likely, the list of 23 structured categories from the extended interview would be adequate.

One insurmountable problem with telephone surveys is the inability to control for bad data. This is best described as informants who are 'playing' or who are uninterested and not thinking about their responses. The line between good data and bad data is quite fuzzy, and frequently not definable. Yet it is always a factor to consider in the interpretation.

In general, both methods were useful and worthy of refinement and replication. The extended interview should be focused on more detailed information about each alter, while the telephone interview should be based on precoded types of social relations.

Knowing

The results of the two experiments verified that there is enough variation in the way that knowing is perceived that researchers must use caution in how they apply the concept. The wide variation in estimates of personal networks is understandable given the myriad ways in which informants use this concept.

For instance, some informants think of how much they know about someone, while others think about how often they see someone. It is easy to imagine a case where we see someone frequently but know little of them, such as a work relation. Conversely, I might see someone infrequently and know a lot about them, such as a family member. The concepts are quite different.

When analyzing the knowing themes by the level of the knowing scale, there is also evidence of variation. For example, family is a concept which is almost exclusive to the highest knowing categories. In contrast, frequency of contact is hardly ever listed as a criterion for the assignment of the highest level of knowing. Defining knowing, or a network tie, according to frequency of contact could have deleterious effects on the results.

What these data suggest is that we must be certain that the researcher and the informant understand one another as to how the concept of knowing is used. Until a more empirically and functionally defined tie identifier can be developed, we are left with increasing the precision of old ones.

The Social Relation

What do informants mean when they claim to know someone? One way to answer this question is to focus on how they determine how well they know an alter. Recall that informants described their use of the various knowing levels.

Someone who is known well is a relative (an ascribed role which may or may not include frequent contact), a friend (an ambiguous tag which can include a cornucopia of behavioral expectations), someone the informant knows a lot about, or someone whom they contact frequently. Someone who is not known well is usually a non-relative whom informants knew little about, who was not a friend, and was seen infrequently.

The various themes expressed fall into three broad categories. One category is the concrete. These are themes such as frequency of contact, knowledge about the person or duration of relation. Unlike other themes, these are potentially measurable in distinct and valid ways. Indirectly, they also imply the ability to predict the behavior of the alter. Presumably, informants who see an alter with some frequency, know a lot about them and have done so over a long period of time can make better predictions of the alter's behavior than of other alters.

Another theme category is a relation attribute These are themes such as closeness or friendship Taken on face value friend may seem to be a distinct social relation But it does not take long to determine that friend is much too broad and ambiguous to make any statement about the relation, other than that it is most likely not adversarial[33] Reading through the descriptions it appears that friendship is not a description of a relation, but an attribute that has degree Like closeness informants used phrases such as "Not much of a friend" or "A very good friend " This is really a quality of the relation that in general, implies the presence or absence of trust or companionship. Unfortunately the researcher must fill in the blanks for the concrete meaning of friendship

Other than a few exceptions (like "Work or School"), family or relative is the only theme that describes a type of relation It is interesting that most anthropologists assume the role of kinship to have decreased with the advent of complex societies This may be true in comparison to Yanomama society but among these informants kin relations are far from trivial In fact this is the only relation type that appears consistently across time and space it is an ascribed relation that is unique

So when an informant says they know someone the data suggest they are referring to these three broad categories of knowing the concrete (i e frequency of contact and amount of personal knowledge), attributes such as closeness and friendship and whether the alter is a relative But who do informants know? What types of social relations exist among their 60 alters?

In terms of demographics, most informants' alters were of the same sex and race although the sex distribution was not highly skewed

---

[33] Even this might not be true in that some friends have very adversarial relationships indeed for some people this is the basis of a good friendship

Similarly, alters tended to be close to the age of informants as well as being met and living in Gainesville[24]. Most alters are demographically similar to informants and most live close by, within Gainesville (40%) or at least in Florida (29%).

Overwhelmingly, the social relations are positive or neutral. Very few informants listed alters where the relation was considered negative. Perhaps related to this is the high percentage of relations that are seen as mutual. Most informants feel that the effort of maintaining the relation, whatever that entails, is approximately equal. Only situational relations account for any other noticeable percentage. But given the role of location, changes in location are very likely to affect the majority of them. This is supported by the large percentage (29.4%) of alters who would not be known after a year given a change in location.

Labeling the relation depends initially on the circumstances which provided for its inception. Subsequent to this initial casting, a relation can become multiplex, but most relations are founded on one identifiable situation. In a sense, all relations are initially situational, but some situations carry more institutional expectations than others. For instance, a social relation based on a hobby is more flexible than one based on work. I have more choice of participating in the hobby relation than the work relation. The family relation is the pinnacle of inflexibility. Although I can choose not to associate with my brother, it is virtually impossible to negate the relation if other family exist to foster the expectations that go with it.

The structured coding results displayed in Table 18 summarized the types of relations these informants listed. Friend is the highest occurring label, but this represents an attribute rather than a relation type. People rarely meet randomly and become friendly. This is an attribute that develops secondary to some other relation.

---

[24] Gainesville was the modal category, but adding Florida, not Gainesville and outside Florida together yields a larger percentage than Gainesville alone.

Similarly, confidant and acquaintance are not descriptive of the foundation of the relation, that is, how it was conceived. As with friendship, they are attributes of a relation. Unlike friendship, they imply more information. Acquaintance suggests that the informant knows little about the alter or sees him/her infrequently. Confidant implies the opposite. One or more of these three attributes are present in over 70% of the alters[35]

The remaining labels, those of any frequency, are truly tags that indicate some basis for the relation to occur. Some of these types can be collapsed since they cannot cooccur. For instance, no alter can be both a close blood relative and a distant blood relative. Nor can an alter be a co-worker and an employee, given my definition. Thus the four broader categories of relatives, work, step relation and school occur in 25.4%, 19.7%, 19.2% and 15.3% of the alters. The remaining types of any consequence are hobby, organization, religious and neighbor relations.

Relatives and work are highly represented. It is not surprising that, given the structure of most informants' daily activities, that most of the people they know are family, which is an ascribed relation, and work, which is also largely an ascribed relation. Perhaps less expected are the number of intermediary ties. A large percentage of informants' networks are made up of relations via another alter. Given the results of the telephone survey, the large percentage for school is likely a sample bias. The remaining types are highly dependent on the interests of the individual and consequently are selectively represented.

---

[35] Note that up to three labels were associated with an alter. This allows for co-occurrences of labels, which frequently happened with friend and confidant.

A full 45.1% of alters are relatives or work relations, since it is probably safe to assume that these are not cooccurring types. Is this a pattern that one would expect to find in Mexican or German culture[36]?

Clique Definition

Having demonstrated some idea as to the types of social relations that informants have, what are informants' perceptions of the structure of these relations? In other words, what subgroups exist among these relations and what are their characteristics?

Using the ADCLUS model on a binary matrix which did not include ego, 19 distinct groups were identified from analyzing the 45 useable matrices. Since 45.1% of all alters are either relatives or work related, it is not surprising that clusters which are founded on these relations occur in 100% and 73.3% of informants, respectively. Less obvious is the subgroup detail that exists in these broader categories.

For instance, work networks come in two types: the current work place and former work place(s). It is almost a truism that informants would have a network centered around their job, or more than one job network. But it is not so obvious that 42.2% of informants would have networks from past jobs that they maintain for any period of time. Some informants have former job networks that have survived for many years, more or less intact. Others do not. Further, college students, who by and large have part time retail jobs, do not tend to form work clusters of an enduring nature.

The detail among subgroups of relatives is more interesting. Again, the existence of maternal, paternal, close, generalized (close and distant) and even significant other's family are intuitive. But the fact that 55.5% of the informants had one or more relative subgroup that included family friends as a highly interactive member is not so obvious.

---

[36] Many of the categories detailed here are culture specific. For instance, could a work relation be defined among the Yanomamo or Kalahari societies where distinct behavioral domains are difficult identify?

Another clique type that was not expected was the "network via." This clique is comprised of a set of two step relations that are highly interactive. A clear example is the family of a coworker or in some cases in-laws. A person with a large proportion of network via cliques demonstrates a fragile network since changes in the lifecycle of relatively few alters can result in the loss of many. For instance, in Betty's case where work relations, and the vias associated with it represent a very large part of her personal network, would realize a complete shift in her personal network. Some people might find this change refreshing.

The correlations which summarize the propensity for clique types to reoccur present mixed results. One of the more interesting is hobby cliques which tend to occur for informants who have general family but tends to not occur with informants who have close family. In fact close family and general family are negatively associated. Thus most informants that have family cliques that contain distant family tend not to have exclusively close family cliques. Hobby groups might then be viewed as a surrogate for close family. This might be pertinent in Steve's case where several close relations exist among alters in hobby groups.

The importance of family in cliques is carried through in the analysis of overlap. Overlapping cliques occurred in the majority of informants. Most are accounted for by sub-cliques within larger cliques. The remainder were, in most cases, an overlap between a family clique usually close family, and some other clique. Depending on how one defines a clique some of these would certainly disappear being absorbed in larger cliques.

How much overlap is expected due to chance is a difficult question to answer. This is a function of the number of clusters, their size as well as the type of social relation which defines it. It is important to point out that the overlaps that were noted were based on strong relations. They are not trivial, ineffectual overlaps. By itself, the high number of overlaps suggests something more than chance operating. Either the informant is a weak tie between otherwise unrelated subgroups, or the informant is enmeshed in networks which, for other reasons, exhibit common social relations.

Conclusion

In the final analysis, these results suggest that network researchers should be conscious of the variability in the way informants perceive subgroups. For example, all informants do not perceive the same family subgroup types as others. The generic term 'family' is a category that includes paternal, maternal, close, general, family friends and a variety of combinations.

I have no doubt, however, that when asked to list the members of "the family network", an informant will comply. The informant assumes the researcher knows what he or she is looking for and will adapt to the task. But is this what the researcher really wants? If one is interested in a network effect, it is more sensible to elicit the membership of cliques rather than members of a category, categories such as "the family network". These results suggest that category and clique membership are frequently not the same.

The results also suggest caution in the use of the word 'know' as a cue. Too often social scientists depend on abstract and ill defined cues to generate data with little regard for the variability caused by the interpretation of the cue itself. To some degree, all cues are subject to variable interpretation, but care can be taken to create a standard definition that will eliminate this as a source of error. It is a fact that informants do interpret knowing in more than one way.

As with all descriptive studies, most of these results are limited to the 47 informants who participated in the in-depth experiment. Only these results derived from the telephone survey are sufficiently representative to be generalized to Floridians as a whole. However it is unlikely that Floridians are the same as Mexicans or Yanomamo in their perceptions of clique structure. A replication of this study or something like it, in other cultures would no doubt, be interesting.

This experiment contributes to the goal of building better experiments. It has been an in depth study of cues and standard experimental assumptions. Hopefully the application of these findings to new network instruments will further our goal of understanding and predicting behavior.

Appendix: Categories of knowing presented to informants

1=2 step relation
2=3 step relation
3=Went/go to school together
4=Coworker/colleague (not boss)
5=Neighbor
6=Relative (blood, close)
7=Relative (blood, distant)
8=Relative (marriage, close)
9=Relative (marriage, distant)
10=Hobby (same hobby)
11=Organization (same organization)
12=Instructor/teacher
13=Student (your student)
14=Boss/supervisor
15=Employee/supervisor
16=Service (supply service)
17=Military
18=Romantically involved
19=Ex (spouse, lover, etc.)
20=Religious affiliation
21=Confidant
22=Friend
23=Acquaintance
24=Other

REFERENCES

Alba Richard D
1973 "A Graph theoretic definition of a sociometric clique " Journal of Mathematical Sociology 3 113 126

Arabie Phipps J Douglas Carroll Wayne DeSarbo and Jerry Wind
1981 "Overlapping clustering A new method for product positioning " Journal of Marketing Research 18 310-317

Athanasiou Robert and Gary Yoshioka
1973 "The spatial character of friendship formation " Environment and Behavior, 5 43 65

Babbie Earl
1989 The practice of social research Fifth Edition Wadsworth Publishing Belmont California

Barnes J A
1954 "Class and committees in a Norwegian island parish " Human Relations 7 39 58

Bernard H R and Peter D Killworth
1979 "Why are there no social physics?" Journal of the Steward Anthropological Society 11 33-58

Bernard H R Peter D Killworth and Christopher McCarty
1982 "Index An informant-defined experiment in social structure " Social Forces 61 99 133

Bernard H R E C Johnson P D Killworth and S Robinson
1987 "Estimating the size of an average personal network and an event subpopulation some empirical results " American Statistical Association Proceedings of the Section on Survey Research Methods

Bernard H R Peter D Killworth Michael J Evans, Christopher McCarty and Gene Ann Shelley
1988 "Studying social relations cross culturally " Ethnology 27 155 179

Bernard H R E C Johnson P D Killworth C McCarty S Robinson and G A Shelley
1989 "Estimating the size of an average personal network and of an event subpopulation " In The Small World M Kochen (ed) Norwood New Jersey Ablex Publishing Corporation

Bock, R D and S Z Husain
1950 "An Adaption of Holzinger's B coefficient for the analysis of sociometric data " Sociometry 13 146 153

Bott Elizabeth
1953 "Urban families Conjugal roles and social networks " Human Relations, 8 343 383

1957 Family and social network London Tavistock

Burt R S
1982 Toward a structural theory of action New York Academic Press

Coleman James S
1961 "Relational analysis " Human Organization, 17 28-36

Comte, Auguste
1896 The positive philosophy Translated and condensed by Harriet Martineau London G Bell

Doreian Patrick
1969 "A note on the detection of cliques in valued graphs " Sociometry 30 237-242

Durkheim, Emile
1933 (original 1893) Division of labor in society Translated by G Simpson New York Macmillan

1938 (original 1895) The rules of sociological method Translated by S Solovay and J Mueller New York Free Press

Epstein A L
1969 "Gossip norms and social network " In Social networks in urban situations J Clyde Mitchell (ed) Manchester Manchester University Press

Everitt Brian
1974 Cluster analysis London Heinemann Educational Books

Festinger Leon
1949 "The analysis of sociograms using matrix algebra " Human Relations 2 133 138

Fischer Claude S
1982 "What do we mean by 'friend'? An inductive study " Social Networks 3 287 306

Forsyth E and L Katz
1946 "A matrix approach to the analysis of sociometric data Preliminary report " Sociometry 9 340-347

Freeman Linton C
1991 "The sociological concept of 'group' An empirical test of two models " In press

Freeman L C and C R Thompson
1989 "Estimating acquaintanceship volume " In The Small World M Kochen (ed) Norwood New Jersey Ablex Publishing Corporation

Granovetter Mark
1973 "The strength of weak ties " American Journal of Sociology 78 1360 1380

Hammer Muriel
1984 "Exploration into the meaning of social network interview data " Journal of Social Networks, 6 341-371

Harary F., R Z Norman and D Cartwright
1965 Structural models An introduction to the theory of directed graphs New York Wiley

Harris Marvin
1968 The rise of anthropological theory A history of theories of culture New York Harper and Row

Holland P W and S Leinhardt
1970 "A method for detecting structure in sociometric data " American Journal of Sociology 76 492-513

Hubbell Charles H
1965 "An input-output approach to clique identification " Sociometry 28 377-399

Jain Anil K and Richard C Dubes
1988 Algorithms for clustering data Englewood Cliffs New Jersey Prentice Hall

Kapferer B
1969 "Norms and the manipulations of relationships in a work context " In Social Networks in Urban Situations J Clyde Mitchell (ed) Manchester Manchester University Press

Killworth Peter and H R Bernard
1974 "CATIJ A new sociometric and its application to a prison living unit " Human Organization 33 335-350

Killworth Peter D H R Bernard and Christopher McCarty
1984 "Measuring patterns of acquaintanceship " Current Anthropology 25 381-397

Kroeber Alfred
1917 "The superorganic " American Anthropologist 19 163-213

Kuhn Thomas S
1962 The structure of scientific revolutions Second Edition Chicago The University of Chicago Press

Lankford Philip M
1974 "Comparative analysis of clique identification methods " Sociometry 37 287-305

Loomis Charles P
1944 "Visiting patterns and Miscegenation at Oxapampa Peru " Rural Sociology 9 68-97

1950 Rural Social Systems New York Prentice Hall

Luce R Duncan
1950 "Connectivity and generalized cliques in sociometric group structure " Psychometrika 15 169-190

Luce R Duncan and Albert D Perry
1949 "A method of matrix analysis of group structure " Psychometrika 14 95-116

Massarik F and P Ratoosh
1965 Mathematical explorations in behavioral science
Homewood, Illinois R D Irwin and Dorsey Press

Mayer P
1961 Townsmen or tribesmen Capetown Oxford University Press

McQuitty Louis L
1957 "Elementary linkage analysis for isolating orthogonal and oblique types and typal relevancies " Educational and Psychological Measurement 17 207 229

Miller, George
1956 "The magical number seven plus or minus two Some limits on our capacity for processing information " Psychological Review 63 81-97

Mitchell J Clyde (ed)
1969 Social Networks and Urban Situations Manchester Manchester University Press

Monge Peter R and Kenneth Kirste
1980 "Measuring proximity in human organization " Social Psychology Quarterly 43 110 115

Moreno J L
1934 Who shall survive? Beacon New York Beacon House Inc

Nadel, S F
1957 The theory of social structure. Glencoe Illinois The Free Press

Peay Edmund R
1974 "Hierarchical clique structures " Sociometry, 37 54 65

1976 "A note concerning the connectivity of social networks." Journal of Mathematical Sociology 4 319 321

Pool, I S and M Kochen
1978 "Contacts and influence " Social Networks, 1 5 51

Popper Karl
1953 "Science conjectures and refutations " in Introductory readings in the philosophy of science E D Klemke Robert Hollinger A David Kline (eds) Buffalo New York Prometheus Books

Radcliffe-Brown A R
1931 "The social organization of Australian tribes." Oceania 1 426 456

1933 (original 1922) The Andaman Islanders Cambridge Cambridge University Press

1940 "On social structure." Journal of the Royal Anthropological Society of Great Britain and Ireland 70 1 12

Rapoport Anatol
1979 "A probabilistic approach to networks " Social Networks, 2 1-18

Riffenburgh Robert H
1966 "A method of sociometric identification on the basis of multiple measurements " Sociometry 29 280-290

Rogers, Everett and D Lawrence Kincaid
1981 Communication networks toward a new paradigm for research New York Free Press

Saris Willem E
1991 Computer-assisted interviewing Newbury Park California Sage Publications

Seidman Stephen B and Brian L Foster
1978 "A graph theoretic generalization of the clique concept " Journal of Mathematical Sociology 6 139 154

Simmel Georg
1955 Conflict and the web of group affiliations Glencoe Illinois Free Press

Strauss David and Linton C Freeman
1989 "Stochastic modeling and the analysis of structural data " in Research methods in social network analysis Linton Freeman Douglas White and A Kimball Romney (eds.) Fairfax, Virginia George Mason University Press

Sudman Seymour
1988 "Experiments in measuring neighbor and relative social networks " Journal of Social Networks 10 93-108

Wellman Barry and S D Berkowitz (eds)
1988 Social structures A network approach Cambridge Cambridge University Press

Wolfe A W
1978 "The rise of network thinking in anthropology " Social Networks 1 53-64

BIOGRAPHICAL SKETCH

Christopher McCarty was born in Parkersburg, West Virginia and lived there until age 18 While pursuing an undergraduate degree in anthropology at West Virginia University he met and collaborated with H Russell Bernard, his mentor in anthropology, and Peter D Killworth a theoretical oceanographer in the area of social network analysis Upon receiving his BA, he entered the University of Florida to pursue a PhD in anthropology

While at the University of Florida, McCarty continued to participate in a fruitful program of study with Bernard and Killworth He also performed fieldwork for his MA in anthropology in an Otomi indian village in Mexico An interest in agricultural economics led to his enrollment in the department of Food and Resource Economics to work towards and MS in that area Over the years McCarty has conducted more fieldwork in Mexico, and participated in a language school in Bulgaria He also held several graduate assistantships and consulting positions performing data analysis and database applications for medical environmental and development agencies.

For the past five years McCarty has worked as a data analyst and researcher at the Bureau of Economic and Business Research at the University of Florida Most recently McCarty has run a telephone survey of Floridian consumers and taught marketing research methods In the future he hopes to become more involved in teaching and consumer research

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy

H. Russell Bernard, Chair
Professor of Anthropology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy

Peter D. Killworth
Professor of Theoretical
Oceanography

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy

Ronald Cohen
Professor of Anthropology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy

Paul Doughty
Professor of Anthropology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy

Richard Lutz
Professor of Marketing

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy

Anthony La Greca
Professor of Sociology

This dissertation was submitted to the Graduate Faculty of the Department of Anthropology in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy

December 1992

Dean, Graduate School